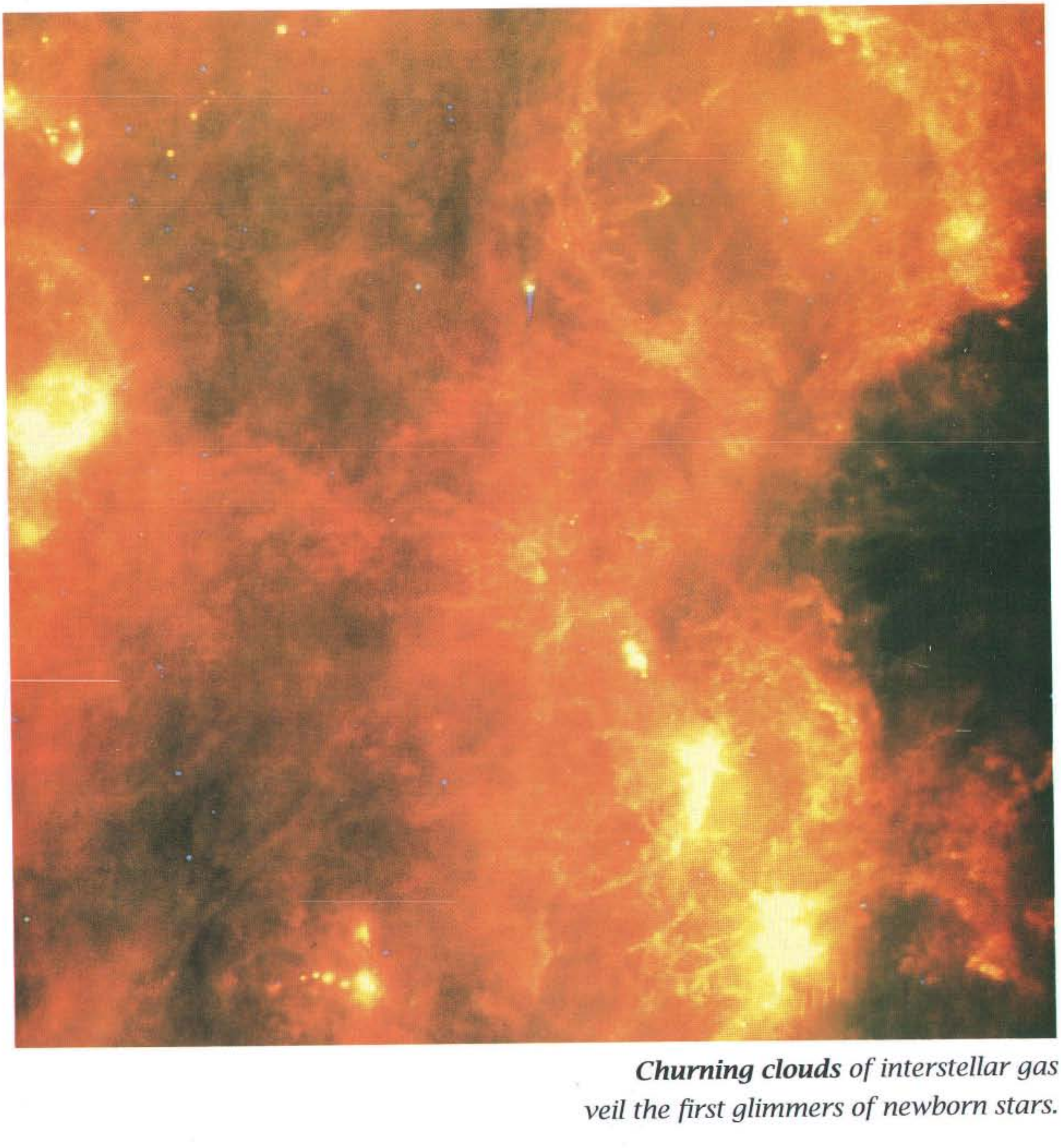
# SCIENTIFIC AMERICAN

JULY 1991
$3.50

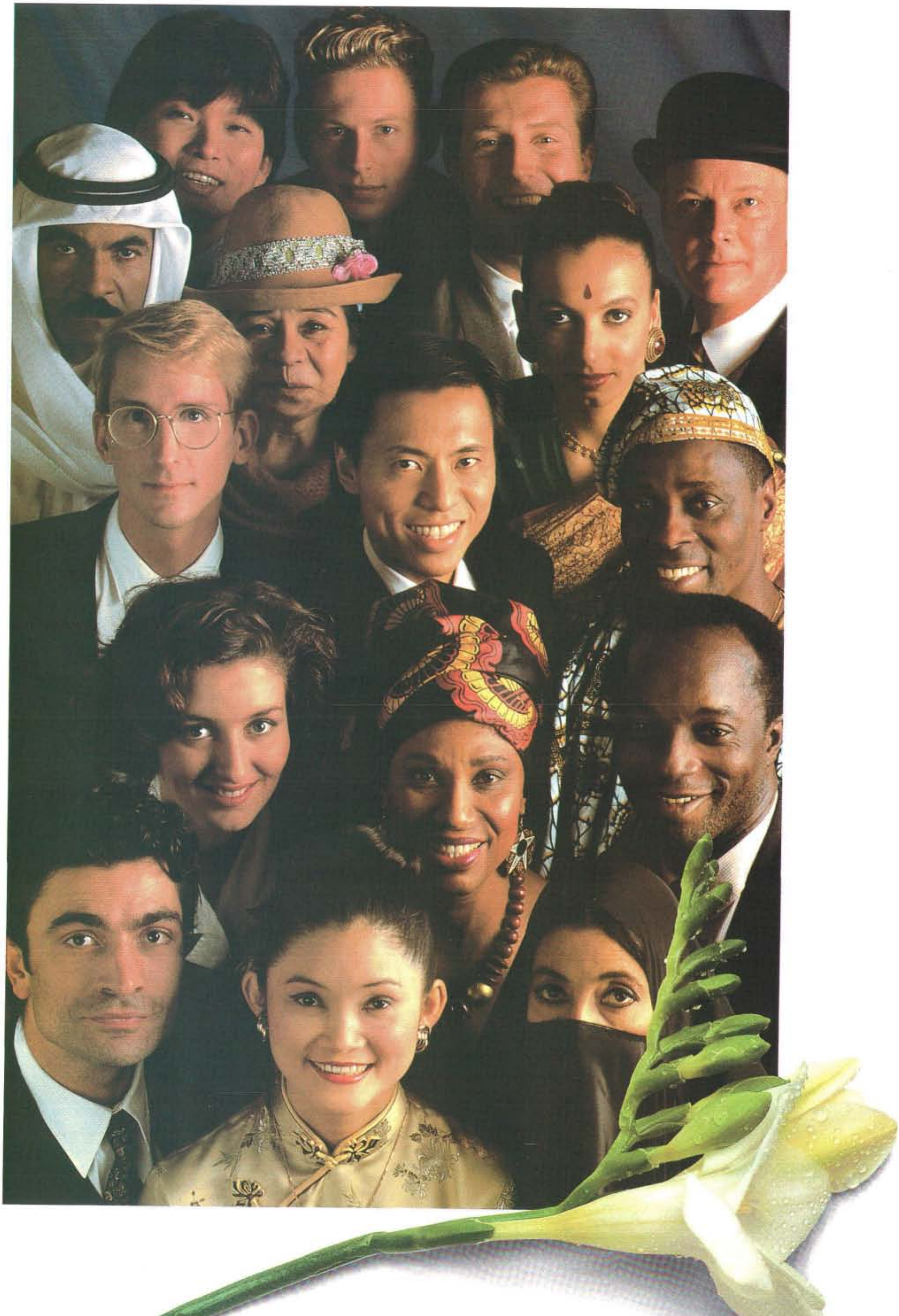
*Hope that strokes may soon be treatable.*

*Do "smart" airplanes have the Right Stuff?*

*When the local druggist sold heroin—legally.*



***Churning clouds** of interstellar gas veil the first glimmers of newborn stars.*



# The art of being local worldwide.

A hundred years of expertise in Power Generation, Transmission and Distribution, Industrial Automation, Transportation, and Environmental Systems makes ABB a world leader in electrical engineering.

That leadership calls for global resource – and local knowledge. Our 215,000 employees serve customers and communities in 140 countries.

If your interests are international, you'll find ABB companies, products and services wherever you do business. If your interests are local, you'll find us right there with you – as insiders.

That's the art of being local worldwide.

ABB Asea Brown Boveri Ltd
Reader Services Center
P.O. Box 822
CH-8021 Zurich, Switzerland

ABB
ASEA BROWN BOVERI

SCIENTIFIC AMERICAN
July 1991 Volume 265 Number 1

**20**

## Opium, Cocaine and Marijuana in American History
*David F. Musto*

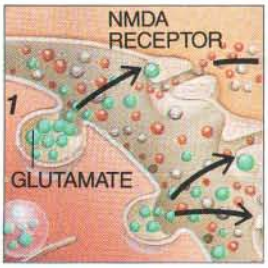
A year before Bayer introduced aspirin in 1899, the drugmaker coined another well-known name for a then popular remedy: heroin. Soon after, the social climate changed, and many mood-altering drugs were made illegal. The author argues that rational drug policy cannot be achieved without keeping such historical reversals in mind.

**28**

## The Early Life of Stars
*Steven W. Stahler*

Our sun experienced a turbulent youth long before reaching maturity as a stable source of fusion energy. Astronomers are now piecing together this complex life cycle of stars. The process begins when clouds of interstellar gas coalesce into protostars discernible only in the infrared and culminates in one of the billions of optically visible stars.

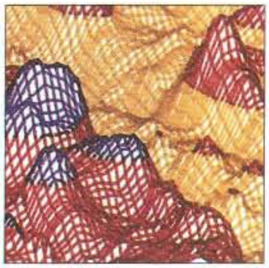
**36**



## Stroke Therapy
*Justin A. Zivin and Dennis W. Choi*

Physicians have long helplessly stood by when strokes felled their patients. But recent insights into mechanisms that destroy nerve cells are leading to treatments that may minimize damage to the brain. Clinical trials of clot-dissolving drugs are showing promise, and tests of others that slow cell death are under way.

**44**



## Optical Interferometry of Surfaces
*Glen M. Robinson, David M. Perry and Richard W. Peterson*

The quality of many products, from photographic film and computer disks to bearings, depends on the microscopic structure of their surfaces. New techniques use computers to analyze optical interference patterns and display surface features as small as a few hydrogen atoms. These methods have already cut costs and improved product performance.

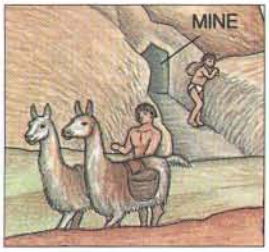
**50**

## Biological Control of Weeds
*Gary A. Strobel*

No one knows for sure when an unwanted plant was first termed a weed, but the idea certainly precedes the writing of the Old Testament. These plants have been pulled up, plowed under and poisoned, but the age-old battle rages on. The latest ploy: enlisting such natural enemies of weeds as insects and fungi. Will these new allies give humans the edge?

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Copyright © 1991 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537.

62



## Copper-Alloy Metallurgy in Ancient Peru

*Izumi Shimada and John F. Merkel*

When Francisco Pizarro and the conquistadores invaded Peru in the 16th century, they carted away tons of gold and silver. Yet they ignored the most ancient and sophisticated metallurgical tradition. For six centuries, copper alloys had been the mainstay of Peruvian technology and commerce.

70

## The Austronesian Dispersal and the Origin of Languages

*Peter Bellwood*

The word, it seems, was borne by farmers seeking new agricultural lands. The ancient diffusion of language through Polynesia is yet another example. Traveling great distances by boat, these societies spread their languages from Taiwan to Madagascar and Hawaii.

76

TRENDS IN TRANSPORTATION

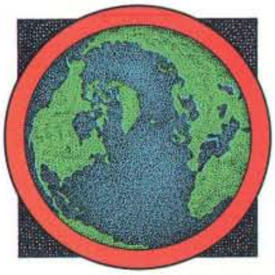
## Along for the Ride?

*Gary Stix*, staff writer

Today's "smart" aircraft can virtually fly themselves to any point on the globe. Computerized navigation systems and flight controls are replacing pilots' Right Stuff with expertise in systems management. Airframe manufacturers insist these are the safest planes ever flown, but pilots are sometimes ill at ease in these "glass cockpits."

# DEPARTMENTS

8



## Science and the Citizen

The far-flung soot from Kuwait's oil fires.... Closing in on a malaria vaccine.... The *Gamma Ray Observatory*.... The biggest black hole.... Quantum effects in microscopic magnets.... A test for antisense RNA.... PROFILE: The Dalai Lama

86

## Science and Business

Teasing light from silicon.... The long wait for ferroelectric semiconductors.... Musical computers that are musician friendly.... Edible holograms.... It's still FORTRAN.... THE ANALYTICAL ECONOMIST: A use for game theory?

5

**Letters**

Dueling linguists.... No guiding theory.... Nonimaging optics.

6

**50 and 100 Years Ago**

1941: Move over silver screen, here comes theater television.

93

**Mathematical Recreations**

A generation of robots that crawl, think and act—just like bugs.

96

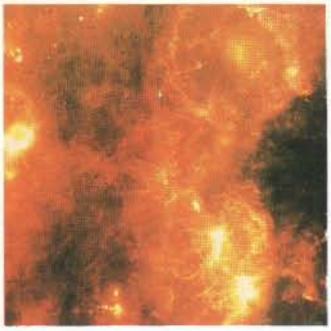
**Books**

A walk on the wild side of prehistory.... Fibbing with maps.

100

**Essay:** *John Timpane*

There's no conflict between the poet and the scientist.



THE COVER photograph is a false-color infrared image that shows various nebulas in the constellation of Orion. In such regions the swirling masses of gas and dust collapse to form stars (see "The Early Life of Stars," by Steven W. Stahler, page 28). The lower right shows the Orion nebula, and the bright patch left of center is the Rosette nebula. Many young stars are visible throughout. The composite photograph, taken by the *Infrared Astronomical Satellite* in 1983, was recently reprocessed to enhance the image and remove flaws.

## THE ILLUSTRATIONS

Cover photograph by Infrared Processing and Analysis Center, NASA/Jet Propulsion Laboratory, California Institute of Technology

| Page | Source |
|---|---|
| 21 | Yale Medical School Library |
| 23 | American Institute of the History of Pharmacy, University of Wisconsin–Madison |
| 24 | Brent Papers, Smithsonian Institution; Randy Santos (*photography*) |
| 25 | Bettmann Archive |
| 26 | Smithsonian Institution; Randy Santos (*photography*) |
| 27 | Time Warner, Inc., © 1981, 1986, reprinted by permission |
| 29 | California Institute of Technology (*top*), National Optical Astronomy Observatories (*bottom*) |
| 30–34 | George Retseck |
| 37 | William J. Powers, Jewish Hospital, St. Louis |
| 38 | Carol Donner |
| 39 | Justin A. Zivin (*top*), Tomo Narashima (*bottom*) |
| 40 | Tomo Narashima |
| 42 | Gary K. Steinberg, Stanford University |
| 43 | Bettmann Archive |
| 44–45 | Glen M. Robinson, David M. Perry and Richard W. Peterson |
| 46–47 | Andrew Christie |
| 48–49 | G. M. Robinson, D. M. Perry and R. W. Peterson |
| 50 | Wilford Hess, Brigham Young University |
| 51 | Gary A. Strobel |
| 54 | Patricia J. Wynne (*top*), Photo Researchers, Inc. (*bottom*), Patricia J. Wynne, with Jerome Rozen, Jr., American Museum of Natural History (*inset*) |
| 56D | Laurie Grace; Patricia J. Wynne (*insets*) |
| 58 | Claude D. Boyette, Southern Weed Science Laboratory (*left*), Gary A. Strobel (*right*) |
| 59 | Laurie Grace; Patricia J. Wynne (*insets*) |
| 60 | Gary A. Strobel |
| 62–63 | Patricia J. Wynne |
| 64 | Joe LeMonnier |
| 65 | Izumi Shimada (*left*), Patricia J. Wynne (*right*) |
| 66–68 | Izumi Shimada |
| 71 | Michael A. Tongg |
| 76–77 | © 1991 by John McGrail |
| 78–79 | Gabor Kiss (*insets*) |
| 80 | Boeing Company |
| 81 | Boeing Company (*top*), Lockheed Corp. (*bottom*) |
| 82 | Armstrong Laboratory, Wright-Patterson Air Force Base |
| 83 | Flight Dynamics, Inc. |
| 84 | Black Star/Jeffrey A. Rycus |
| 93 | Bruce Frisch |
| 94 | Michael Goodman |
| 95 | Michael Goodman (*top*), Bruce Frisch (*bottom*) |
| 100 | Poetry excerpts: from *The Collected Poems of Wallace Stevens*, by permission of Alfred A. Knopf, Inc., and Faber and Faber Ltd; from *The Collected Poems of William Carlos Williams*, Vol. 1: *1909–1939*, by permission of New Directions Publishing Corp.; "Corsons Inlet," by A. R. Ammons, by permission of W. W. Norton; from *The Expectations of Light*, by Pattiann Rogers, Princeton University Press |

# SCIENTIFIC AMERICAN®

*Established 1845*

**EDITOR:** Jonathan Piel

**BOARD OF EDITORS:** Alan Hall, *Executive Editor;* Michelle Press, *Managing Editor;* Timothy M. Beardsley; Elizabeth Corcoran; Deborah Erickson; Marguerite Holloway; John Horgan; Philip Morrison, *Book Editor;* Corey S. Powell; John Rennie; Philip E. Ross; Ricki L. Rusting; Russell Ruthen; Gary Stix; Paul Wallich; Philip M. Yam

**ART:** Joan Starwood, *Art Director;* Edward Bell, *Associate Art Director, Graphics Systems;* Nisa Geller, *Photos;* Johnny Johnson

**COPY:** Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Jonathan Goodman; Daniel C. Schlenoff

**PRODUCTION:** Richard Sasso, *Vice President Production & Distribution; Managers:* Carol Albert, *Prepress;* Tanya DeSilva, *Projects;* Carol Eisler, *Manufacturing & Distribution;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems;* Leo J. Petruzzi, *Manufacturing & Makeup;* William Sherman, *Quality Assurance;* Carl Cherebin

**CIRCULATION:** Lorraine Leib Terlecki, *Circulation Director;* Cary Zel, *Circulation Manager;* Rosa Davis, *Fulfillment Manager;* Katherine Robold, *Assistant Business Manager*

**ADVERTISING:** Robert F. Gregory, *Advertising Director.* OFFICES: NEW YORK: William Buchanan; Peter Fisch; Michelle Larsen; Meryle Lowenthal; Elizabeth Ryan. CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager;* William F. Moore. WEST COAST: 1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025; Kate Dobson, *Advertising Manager;* Joan Berend, Lianne Bloomer, San Francisco. CANADA: Fenn Company, Inc. DALLAS: Griffith Group.

**ADVERTISING SERVICES:** Laura Salant, *Sales Services Director;* Diane Schube, *Promotion Manager;* Mary Sadlier, *Research Manager;* Ethel D. Little, *Advertising Coordinator*

**INTERNATIONAL:** EUROPE: Roy Edwards, *International Advertising Manager,* London; GWP, Düsseldorf. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.

**ADMINISTRATION:** John J. Moeling, Jr., *Publisher;* Marie D'Alessandro, *Business Manager*

**SCIENTIFIC AMERICAN, INC.**
415 Madison Avenue
New York, NY 10017
(212) 754-0550

**PRESIDENT AND CHIEF EXECUTIVE OFFICER:** Claus-Gerhard Firchow

**CORPORATE OFFICERS:** *Executive Vice President and Chief Financial Officer,* R. Vincent Barger; *Senior Vice President,* Linda Chaput; *Vice Presidents:* Jonathan Piel, John J. Moeling, Jr.

**CHAIRMAN OF THE BOARD:** Dr. Pierre Gerckens

**CHAIRMAN EMERITUS:** Gerard Piel

# LETTERS TO THE EDITORS

## Mind Your Language

Although Philip E. Ross correctly portrayed us in "Hard Words" [Scientific American, April] as taking opposite sides on many issues, we would like to join in commending the intelligence, the vivacity and the accuracy of his reporting. We also agree that some of the speculations about linguistic prehistory, which he reported with admirable evenhandedness, have already received sufficient scrutiny and been found wanting.

We feel challenged by Ross to sort out the methodological and factual issues arising from the Nostratic and Sino-Caucasian hypotheses and to initiate the kind of careful evaluation that these theories deserve but have not had. If forced to wager on the outcome, we would still place our bets differently, but we have decided jointly to pursue this inquiry wherever it may lead.

Eric P. Hamp
Department of Linguistics
University of Chicago

Alexis Manaster Ramer
Computer Science Department
Wayne State University
Detroit, Mich.

The quotes attributed to me by Ross have been so severely divorced from their context that they are utterly incoherent to anyone even passingly acquainted with the fundamental principles of linguistic investigation. First, none of my comments refer to the work of Derek Bickerton. They refer to the misinterpretation of his work that I feared would emerge in your pages. Second, "that all [Creole languages] share structural features that reflect innate patterns of mind" is a core principle of modern linguistics: all human languages share such features. There is not a hint of racism in such a notion.

Third, my references to "children under the age of two" and the "unity of humankind" were specific reactions to Ross's rendering of Bickerton's theories. He put them to me as follows: children before the age of two, apes like Washoe, adults when extremely agitated and Creole languages preserve structural features of a "pre-language." It was this reading of Bickerton that I labeled as potentially racist, or open to exploitation by racists, and contrary to the tenets of linguistics.

It seems clear that your editorial staff lacks sufficient grasp of linguistics to understand its own interviews.

Mark R. Hale
Department of Linguistics
Harvard University

## An Immune Reaction

The late Richard Gershon was a vigorous and effective proponent of the importance of suppression in the regulation of the immune system, but in "The Body against Itself" [Scientific American, December 1990], John Rennie gives him too much credit for suppression theory. The statement that when Gershon died in 1983 the field was left "without a theory for inspiration" is inaccurate. I published an idiotypic network theory in the *European Journal of Immunology* in 1975 that included an explicit mechanism for *T* cell-mediated suppression. It was alive and well when Gershon died and continues to be so today. Michael D. Grant, Tracy A. Kion and I are publishing a paper (*Proceedings of the National Academy of Sciences,* Vol. 88; April 15, 1991) in which we show that the theory leads to a new perspective on AIDS pathogenesis.

Geoffrey W. Hoffmann
Department of Microbiology
University of British
Columbia, Vancouver

***Rennie responds:***

The opinion that research on suppression has lacked a detailed guiding theory is not my own: suppression researchers voiced it frequently.

## Illuminating Origins

Early in "Nonimaging Optics" [Scientific American, March], Roland Winston writes, "Nonimaging optics was born in the mid-1960s, when [we] designed the first compound parabolic concentrators." In fact, the use of a nonimaging sapphire cone concentrator to intensify the pumping of a laser (including the square of the refractive index effect) was published in *Applied Optics* (Vol. 1, No. 2; March 1962) several years earlier. Willard S. Boyle and I used these ideas to make the first continuously operating ruby laser pumped by a mercury arc lamp. We also pointed out the importance of the cone concentrator for solar pumping of a laser.

We thought of our ideas as novel at the time. Indeed, we received a patent for end pumping of a laser in general and with a conical nonimaging concentrator in particular. We knew, however, that we were not inventing nonimaging optics, and we cited a 1952 publication by Donald E. Williamson ("Cone Channel Condenser Optics," *Journal of the Optical Society of America,* Vol. 42, No. 10) that described the procedure for designing a cone concentrator.

Donald F. Nelson
Department of Physics
Worcester Polytechnic Institute
Worcester, Mass.

***Winston responds:***

Williamson's work is certainly important, which is why Walter Welford and I mentioned it in our book *High Collection Nonimaging Optics* (Academic Press, 1989). We go on to point out, however, that a simple cone concentrator falls short of the concentration limit attainable with a compound parabolic concentrator. Interested readers can learn more about Williamson's early work through the references listed in my article's further reading box.

### ERRATUM

In "Peering Inward," by Corey S. Powell ["Trends in Geophysics," Scientific American, June], the figure on page 78 appeared with the wrong caption. Readers can obtain a corrected reprint of the article by contacting the editorial offices of Scientific American. The correct caption follows:

"Molten iron outer core circulates vigorously, generating the geomagnetic field. The top map shows the magnetic field at the surface of the core. Blues indicate inward (north) magnetic flux; reds and oranges represent outward (south) flux. The bottom map depicts the inferred motions in the core responsible for the irregular distribution of the field. Arrows show direction and speed (the fastest flows move 30 kilometers per year). Orange indicates upwelling; blue, downwelling."

# At Procter & Gamble, Drives Our Business, Powered

For 150 years, Procter & Gamble has been a technological leader among the leaders in industry worldwide.

P&G employs more than 3000 graduate scientists and engineers (including more than 1000 with PhDs) in 32 R&D facilities in 19 countries, and supports them with over $700 million in R&D spending each year.

These men and women are generating world-class innovations to maintain the growth of P&G's worldwide business. Recent innovations include:

- Inventing a new way to decaffeinate coffee by using an ester which occurs naturally in many fruits, producing a better tasting beverage at substantially lower cost.
- Developing ultra-absorbent hydrophilic gels to make diapers that are thinner to reduce waste, more comfortable, more absorbent and fit better.
- Developing beverages which deliver bio-available calcium and pharmaceutical products which inhibit calcium loss and maintain bone strength in order to address major health problems such as calcium deficiency and osteoporosis.
- Advancing the leading edge of polymer science by developing a unique family of biodegradable polymers which can be formulated into liquid detergents and fabric softeners, providing substantial improvements in cleaning and stain prevention.

The person who is generally recognized as the most productive technologist in Procter & Gamble history is Victor Mills. An engineer's engineer, his technology achievements during a career lasting from 1926 to 1961 were truly remarkable. They included pioneering work on continuous hydrolyzer processes; controlled crystallization processes which remain state-of-the-art today; continuous hydrogenation processes; cake mix milling processes which dramatically improved the quality and acceptance of prepared baking mixes; spray drying processes; and airtight canister packaging. Many products, especially in the food and paper areas, owe their success largely to his creative contributions. He is the holder of 25 patents. He developed processes which improved consumer products to make life better for innumerable millions of people worldwide.

---

In May, 1990 Procter & Gamble announced the formation of The Victor Mills Society to honor those very special technologists in the Company who, through their technology innovations, have made sustaining and important business contributions throughout their careers.

Twelve charter members of the Society were inducted in November, 1990. These individuals have produced major innovations leading to important business benefits. Their inventions have resulted in 166 U.S. patents, with corresponding patent protection worldwide. They have used technology to establish new standards of excellence in fields as diverse as dental care, environmental protection, detergency and bone metabolism. They are recognized by their peers as leaders in their fields.

They provide an unequivocal standard of achievement in technology development to which other technologists can aspire. P&G salutes the charter members of The Victor Mills Society, along with every member of the Company's research and development team. Their efforts, their dedication and their accomplishments are a source of pride to P&G, and to the technological community overall.

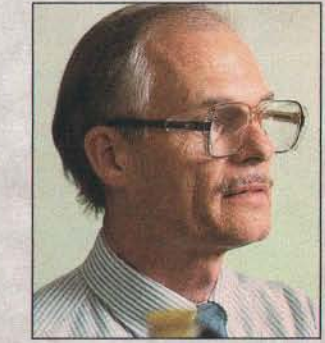
# Innovative Technology By Innovative Technologists!

## The Charter Members of The Victor Mills Society… Honoring Excellence in Technology at Procter & Gamble.

Raymond E. Bolich, Jr.
BS, Chemical Engineering
Beauty Care Division
Cincinnati, Ohio

Kenneth B. Buell
MS, Mechanical Engineering
Paper Products Division
Cincinnati, Ohio

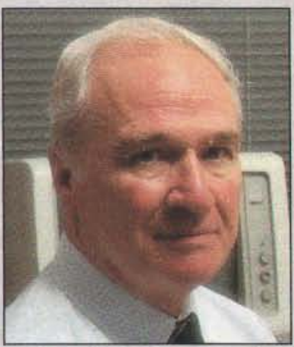
Marion David Francis
PhD, Biochemistry
Norwich Division
Norwich, New York



Eugene P. Gosselink
PhD, Organic Chemistry
Fabric & Hard Surface
Technology Division
Cincinnati, Ohio

Frederick E. Hardy
PhD, Chemistry
Laundry Detergent Products–
Europe
Newcastle, United Kingdom

William D. McKinney
BS, Chemical Engineering
Cellulose & Specialities Division
Memphis, Tennessee



John J. Parran, Jr.
MS, Chemical Engineering
Health & Personal Care
Technology Division
Cincinnati, Ohio

Pedro A. Rodriguez
PhD, Analytical Chemistry
Corporate Research Division
Cincinnati, Ohio

Paul Seiden
BS, Chemical Engineering
Food & Beverage
Technology Division
Cincinnati, Ohio

Geoffrey Stanley
MS, Chemistry
Bar Soap & Household
Cleaning Products–Europe
Newcastle, United Kingdom

Paul D. Trokhan
MS, Chemical Engineering
Paper Products Division
Cincinnati, Ohio

Horst G. P. Wienecke
PhD, Pharmaceutical Chemistry
Health & Beauty Care–Europe
Gross Gerau, Germany

# 50 AND 100 YEARS AGO

## JULY 1941

"We need pilots, pilots, and more pilots; but whereas the mechanically trained youth of America provide us with an abundance of potential mechanized fighting men, we have no pool of flying youths on which to draw for pilots. There is a practical, economical, and effective solution to this problem. We can open aviation to every interested youngster by giving him the opportunity to learn to soar. When he has learned to operate his sailplane, to use the air currents to maintain flight, to meet the ever-recurring small emergencies that come on every glider flight, he will know so much about flying that he can be turned into a power-plane pilot in a fraction of the time and at a fraction of the cost for training a raw beginner. Moreover, he will learn things about the air that a power pilot may never learn until it is too late."

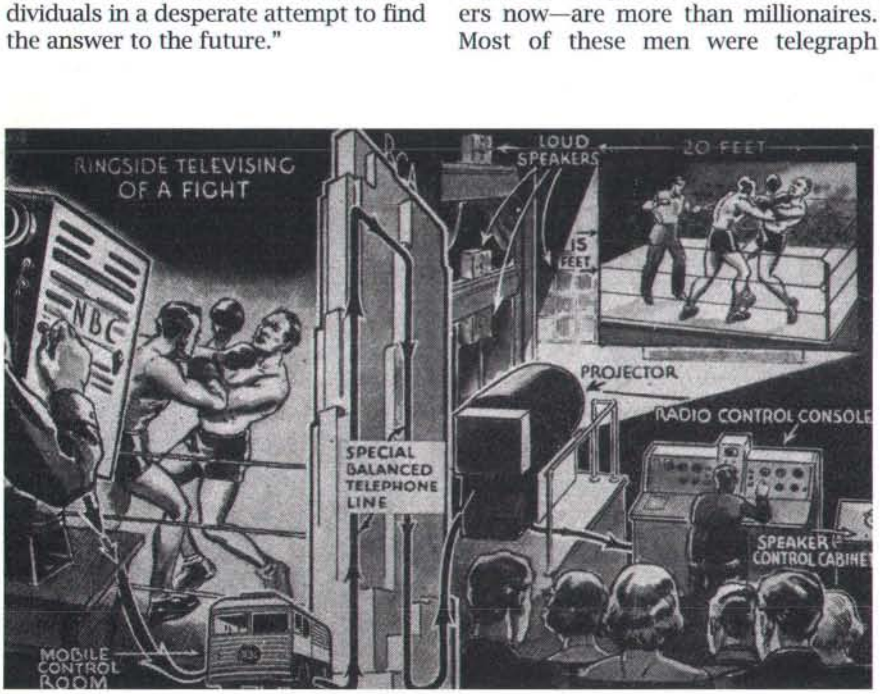
"Mental distress and uncertainty of all peoples has been enhanced by the world's war-like attitude, by international economic and social unrest, with the result that man's age-old query of what lies ahead next week, next year, brings about a dangerous public trend toward futility, and a tendency to turn to so-called mystic rites of cults or individuals in a desperate attempt to find the answer to the future."

"Large-screen pictures constitute the most dramatic development of television in recent days. At an RCA and NBC demonstration in a New York theater, a complete program was presented utilizing equipment that had been specially installed for the purpose [*see illustration for placement of the equipment*]. The program consisted of a playlet put on by 'live' actors, a news broadcast, a newsreel picked up from film, and a ring-side pickup of a championship boxing match. Spectators stated that the quality of reproduction was good, and that the views of the fight were better than could be had from a ring-side seat."

SCIENTIFIC AMERICAN

## JULY 1891

"No scientific body in the country has so many millionaires as the American Institute of Electrical Engineers. At the top of the list is Alexander Graham Bell, whose profits on the telephone are represented by eight figures. Next comes Edison with a seven figure fortune. Brush, of electric light fame, and Elihu Thomson—whose financial future is perhaps brighter than any of the others now—are more than millionaires. Most of these men were telegraph operators, and most of them began their experimenting and study without a dollar."



***Theater television***

"The addition of a small percentage of uranium to steel increases its elasticity, and at the same time its hardness, to an extent that makes its use in the manufacture of guns, armor plates, etc., most desirable, but the scarcity of the material and especially the great difficulty in reducing the ore to metal makes the price of uranium steel too high."

"Mr. C. Vernon Boys has been making measurements of the heat of the moon by means of his very delicate radiomicrometer. His method was to focus the rays of the moon on the face of the radiomicrometer by a reflecting telescope of 16 inches aperture. In the case of a new moon, he found that the heat coming from its disk diminished as you passed from the convex to the concave edge, and that from the dark surface was so slight as not to affect the apparatus."

"*To the Editor of the Scientific American:* Referring to the article in your issue of this date, under head 'Lead Pipe Pierced by Insect,' I have in my possession a lead bullet, that I cut out of the tree under which Grant and Pemberton arranged for the surrender of Vicksburg, Miss. The bullet was lodged just under the bark in the sappy portion of the tree, and has three holes pierced through it by some kind of an insect. One of the holes contained one of these insects at the time I secured the bullet. These facts can be substantiated by two witnesses now living.—Wm. E. Selleck, Chicago, June 13, 1891."

"It is believed that there are traces in the animal kingdom of a law that fixes the extreme duration of life at five times that of growth. This latter period in man may be said to average 21 years. Hence the full span of a perfectly healthy man's life should range from 100 to 105 years. As, however, none are born perfectly free from taint, the expectation of life varies greatly. Every human being starts on his life's journey with a certain life-force; or, in other words, like a clock, he is constructed to run a certain time under given conditions."

# ALL SIMILARITIES ARE PURELY INTENTIONAL.

**There are many similarities between the tires we build for the race track and those we build for your car.**

This is no accident. In fact, we've even given this similarity a name: Technology Transfer.

At Goodyear, it's one of the principal reasons we race. Because when we master the forces involved in keeping a race tire performing at its peak at over 300 kilometres per hour, it's hardly surprising that we advance tire technology in general.

TECHNOLOGY TRANSFER

Eagle Racing Tire

Eagle Street Tire

Advances you can feel in the grip of our road tires. Advances that improve your car's performance. Advances that help make you a better driver.

GOODYEAR

# SCIENCE AND THE CITIZEN

## Burning Questions

*Scientists launch studies of Kuwait's oil fires*

Scientists are mounting a massive worldwide study of the effects of hundreds of burning oil wells in Kuwait. The conflagration, which was ignited by Iraqi troops in late February and is expected to last for at least another year, is obviously important in its own right: even optimists say it may be the worst man-made atmospheric pollution event in history.

The smoke's exaggerated proportions should also provide insights into subtle processes underlying global warming, ozone depletion, acid rain and other controversial atmospheric phenomena. "It's a tragedy for the region, but it's extremely important for science," says Richard D. Small, an atmospheric scientist for the Pacific-Sierra Research Corporation in Los Angeles.

By late May, clear-cut information on the fires remained extraordinarily scarce. "There is a very great lack of atmospheric measurements," says Rumen D. Bojkov, chief of the environmental division at the World Meteorological Organization (WMO) in Geneva, which at the end of April hosted the first scientific conference on Kuwait's fires.

Yet scientists are beginning to form a picture—albeit incomplete and even contradictory—of the fires' impact. At least a half dozen groups—three from the U.S. and one each from Germany, Great Britain and Canada—have completed computer models of the fires' atmospheric effects. Although the modelers made different assumptions about such crucial issues as the rate of oil burning, they all agree that the fires will not have a significant—that is, exceeding the bounds of natural variability—influence on the global climate.

The fires' carbon dioxide output is too small and temporary to contribute to the greenhouse effect and hasten global warming, explains Keith A. Browning, who supervised modeling by the British Meteorological Office. More important, Browning notes, the models suggest that soot cannot rise high enough and stay airborne long enough to dim the sun and cause cooling worldwide, as some scientists, notably Paul J. Crutzen of the Max Planck Institute for Chemistry in Mainz, Germany, speculated before the war.

Measurements made by the Meteorological Office in March seemed to support this view. After 57 hours of airborne observations as far as 1,000 kilometers downwind from Kuwait, the British team concluded that smoke generally rose no higher than five kilometers.

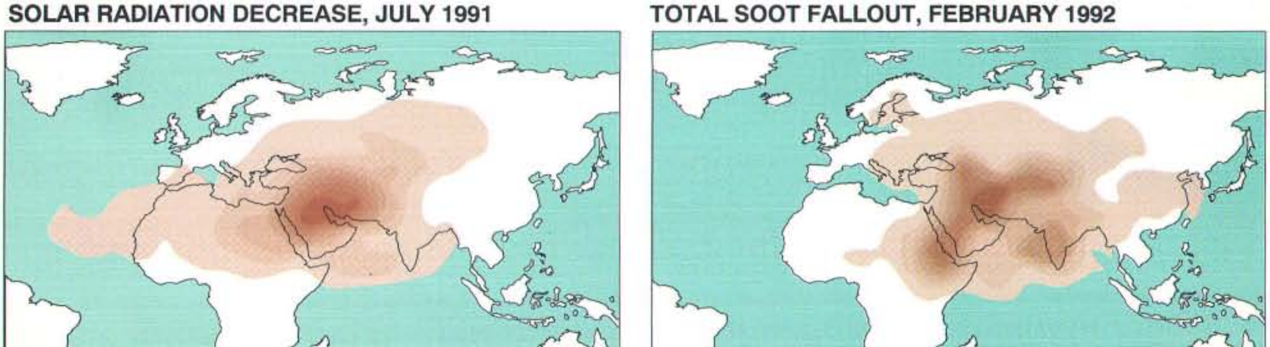
Yet data from more distant regions suggest that some smoke may rise higher and stay aloft longer than the models and the British data indicate. Since February, investigators from the National Oceanic and Atmospheric Administration (NOAA) have recorded numerous "spikes" of soot with an air-sampling instrument at the Mauna Loa Observatory, which sits atop a 4,000-meter mountain in Hawaii. Records of wind patterns show that the sooty air passed over the Gulf region from seven to 10 days earlier.

A similar measurement at Mauna Loa by Robert A. Eldred of the University of California at Davis found levels of soot in March five times higher than during the previous three Marches. Investigators from Nevada's Desert Research Institute are trying to determine whether the soot gathered at Mauna Loa matches samples from Kuwait.

Two sets of balloon observations by researchers at the University of Wyoming suggest that some smoke may have passed over the continental U.S. James M. Rosen found that particle concentrations in the troposphere above Wyoming increased by at least a factor of 10 from early February to late March. Terry L. Deshler independently found particles at up to 100 times their normal levels in a layer of the upper troposphere ranging from 6,000 to 11,000 meters.

The NOAA and Wyoming researchers point out that the suspicious aerosols could come from Asia and that the concentrations observed so far do not seem large enough to affect weather patterns. But if the tropospheric con-

**Computer Modeling Predicts Effects of Kuwait's Oil Fires**


SOLAR RADIATION DECREASE, JULY 1991
0
130
Watts per square meter
TOTAL SOOT FALLOUT, FEBRUARY 1992
0
500
Milligrams per square meter

*SOURCE: Max Planck Institute for Meteorology. Model assumes burning rate of three million barrels a day, beginning February 15, 1991.*

centrations continue to increase or to leak into the stratosphere—where they could stay for years without being washed out by weather—they could begin to have an impact, Deshler says.

Moreover, a global simulation by a group at the Max Planck Institute for Meteorology in Hamburg does not offer a large margin for comfort. During one of its runs, the group erroneously raised its estimate of the rate at which oil is burning in Kuwait to 15 million barrels a day, three times the estimate set forth in May by the WMO. The result, according to Hartmut Grassl, was highly significant global effects, including extreme highs and lows in both temperature and precipitation and unusually frequent storms.

Grassl says this calculation was only briefly mentioned in a paper to be published in *Nature,* but he acknowledges that he and his colleagues were troubled by it. If Saudi Arabian oil fields had also been ignited during the war, he says, the scenario could "easily" have come to pass. The calculation also suggests that the catastrophic scenarios of Crutzen and others were "not as far out in left field as everyone thought," notes Owen B. Toon, a climatologist at the NASA Ames Research Center.

The modelers also examined the possibility that the smoke might affect the Asian monsoon, a rainy season crucial to agriculture in southern Asia. The monsoon occurs when warm air rises over the continent in the summer months, pulling moist air in from the oceans. Before the war, several scientists warned that the smoke might reduce thermal convection over the continent and thereby cause a drought.

So far all the models have contradicted this scenario, according to Jean-Pierre Blanchet, a modeler from the Atmospheric Environment Service in Downsview, Ontario. The models indicate that as the smoke disperses over the Tibetan plateau, it will sink lower to the ground, and the heat it emits into the lower atmosphere will begin to exceed its cooling influence. The net impact, Blanchet says, is a slight warming of the atmosphere and thus, possibly, an enhancement of the monsoon.

Indeed, the Max Planck model predicts that the monsoon might arrive earlier and more forcefully than usual. This statement, Blanchet says, raises a disturbing question: Could the smoke from Kuwait have helped trigger the huge typhoon that struck the coastal region of Bangladesh on May 1, killing more than 100,000 people? Grassl thinks not. He notes that typhoons commonly strike in early May.

But Tiruvalam Krishnamurti of Flor-

## Why Are Data from Kuwait Being Withheld?

In late February the Bush administration hurriedly convened a task force to assess the effects of the hundreds of oil wells burning in Kuwait. Researchers from the task force, which is headed by the Environmental Protection Agency, have been monitoring smoke from the fires continuously since early March.

By May, however, the EPA had issued only a single report, whose main conclusion was that emissions in the smoke were not "at levels of concern." Moreover, EPA officials turned down press requests for interviews with task-force researchers. "They're over there to do monitoring," explained Mary J. Mears, an EPA spokesperson. "We don't want them talking to the press."

Scientists studying Kuwait's fires from afar have also been silenced. In early April, Joyce E. Penner of the Department of Energy's Lawrence Livermore National Laboratory was told by her supervisors not to present a computer simulation of the fires at a scientific conference in Vienna. In May other DOE researchers said they were still under orders not to discuss the fires without clearance from DOE headquarters.

Kuwait and Saudi Arabia may be responsible for some censorship, according to a congressional staff member who has been quietly looking into the issue. Both governments, he said, are concerned that reports about the fires' effects will delay the return of citizens who have fled the region. Mears confirmed that the task force is "supposed to give information to Kuwait and Saudi Arabia first" before considering it for release.

Those countries' sensitivities, however, cannot explain the EPA's suppression of observations of elevated soot levels at a National Oceanic and Atmospheric Administration (NOAA) observatory at Mauna Loa, Hawaii. By early April, NOAA researchers in Boulder, Colo., had tentatively concluded that they were detecting soot from Kuwait. After senior scientist David J. Hofmann gave his approval, NOAA's public affairs officer in Boulder, William J. Brennan, prepared a press release about the observations.

When the release was passed on to John J. Kasper, who heads press relations for both the EPA and the interagency task force, he ordered it withheld. "We were shocked," said a NOAA scientist, who requested anonymity. "That had never happened to us before." NOAA finally issued a press release on May 1 following inquiries about the Mauna Loa soot from SCIENTIFIC AMERICAN.

Kasper later defended the need for information controls, declaring that "you've got to have someone in charge of press releases and announcements, or the whole thing could get very confusing." He characterized the Mauna Loa release as "cheap headline grabbing" and denied that his decision was intended to conceal potentially embarrassing information.

Still, some scientists think the political implications of the findings played a role in the decision. Lara A. Gundel, an aerosol specialist at the Lawrence Berkeley Laboratory, points out that the first suspicious "spike" of soot at Mauna Loa arrived in early February, well before the Iraqis began torching Kuwait's oil wells. The timing suggests, she says, that the soot may have resulted from Allied bombing of Iraqi oil refineries and storage tanks.

Other data support the conclusion that Allied bombing caused significant pollution. A document prepared in early March by the Defense Nuclear Agency and obtained by the Natural Resources Defense Council (NRDC) through a Freedom of Information request notes that Iraq's neighbor Iran experienced "repeated black rain events starting Jan. 22." Satellite images made in mid-February by Landsat-5 and NOAA-11 reveal smoke plumes several hundred kilometers long emanating from various regions of Iraq.

The U.S. may have another, rather subtle motive for wanting to limit revelations about the war's environmental consequences, notes William M. Arkin, a national security expert who is writing a book on the war. Officials in the State Department and the Pentagon, he says, have privately told him they are concerned such revelations will spur demands that Saddam Hussein be tried for environmental war crimes. These sources suspect Hussein would defend himself under "military necessity" exclusions: smoke from burning oil wells, for example, concealed Iraqi troops from Allied bombers. The officials worry that this defense might then lead to calls for stricter environmental protection provisions in the international rules of war, Arkin explains. Spokespersons for the Pentagon and State Department declined to comment. —*John Horgan* ida State University, an expert on the monsoon, adds that the Bangladesh storm was accompanied by unusually severe flooding—two feet higher than ever previously recorded, according to one report. The widespread rains that followed the typhoon, which have hampered relief efforts, are also unusual, he says. Krishnamurti plans to examine the possible linkage between these phenomena and Kuwait's fires.

As for areas nearer Kuwait, the models differ in their details but are still in general agreement, according to Joyce E. Penner, a modeler at Lawrence Livermore National Laboratory. The models predict drops in surface temperatures caused by the smoke's sun-dimming effects of up to 10 degrees Celsius in a region several hundred kilometers around Kuwait and drops of a degree or two in a region as large as 1,000 kilometers in radius.

Acid rain, a by-product of sulfur dioxide, should affect regions some 2,000 kilometers from Kuwait and possibly as far away as China, Penner says. The British model also predicts that soot will fall on snowy regions, causing rapid melting and flooding. All these phenomena might ruin crops and cause other damage.

Meteorological data from Iraq, Iran, Pakistan and other countries likely to be affected are still scarce. But at the WMO meeting in late April, Soviet scientists announced that unprecedented levels of acid rain had been falling in southern Russia. Satellite images have also shown smoke and darkened snow in Pakistan and northern India.

In Kuwait itself, the sky and land are dark, and temperatures are unseasonably low. A report issued by the Environmental Protection Agency in May said pollutants were not "at levels of concern," but other observers presented a different view. The National Toxics Campaign, a private environmental group in Boston, reported finding high levels of cadmium, lead and other toxins in the smoke. An April 30 Reuters story noted a "dramatic" rise in hospital admissions for asthma in the Kuwaiti town of Ahmadi.

In addition, Kuwait is plagued by oil "lakes," some several kilometers across and more than a meter deep, created by oil spilling from unlit wells. To combat this problem, firefighters have been forced to ignite deliberately some wells that cannot be quickly stanched. Pockets of volatile gases can build up under the lakes' crusty surface.

In early May, five people traveling in three vehicles were killed when they crossed a pool of oil and ignited underlying gases, according to Henry W. Kendall, a physicist at the Massachusetts Institute of Technology, who is advising the Kuwait Petroleum Companies. Unexploded U.S. cluster bomblets and Iraqi mines also pose a threat.

In spite of these difficulties, firefighters managed to cap more than 100 damaged wells by late May, but more than 400 were still blazing. Others still set the figure above 500. Kuwaiti officials, Kendall notes, hope to put out half of the fires by this fall and two thirds by next spring. But he adds that putting out the final "recalcitrant" fires could take much longer.

In the meantime, scientists are rushing to the region to gather more data. In mid-May, a rickety old Convair from the University of Washington and a somewhat newer Electra belonging to the National Center for Atmospheric Research (NCAR) flew to Bahrain for a month of atmospheric studies. These missions, says Darrel Baumgardner, a member of the NCAR team, should resolve such crucial issues as the size, shape and light-absorbing properties of the soot particles and their "wettability," or susceptibility to being washed out in rain. Without such data, "it's premature to rule out global effects," Baumgardner contends.

The Defense Nuclear Agency is also planning an expedition for later this summer, according to David L. Auton, chief of the agency's radiation policy division. The trip's purpose, Auton says, is to gather information that will help U.S. armed forces cope with oil fires in future wars. *—John Horgan*

## Proteins 2, Malaria 0

*Malaria-free mice offer clues for developing a human vaccine*

Eighteen mice in Bethesda, Md., may have just carried science closer to one of the holy grails of infectious disease medicine, a human malaria vaccine. Then again, they may not have. Stephen L. Hoffman, director of the malaria program at the Naval Medical Research Institute, and his colleagues reported recently in *Science* that a novel vaccine involving two malarial proteins had completely protected the mice from infections. What remains to be seen, however, is whether such success in mice will translate into help for humanity.

A century ago a British colonial in the tropics could often stave off a bout of malaria with a few doses of quinine water. The two billion people at risk today have no such easy remedy. Chloroquine, mefloquine and other drugs that have been used to treat malaria are losing their potency as resistant strains evolve and spread throughout the world. Each year about 110 million new cases are diagnosed and the disease kills between one and two million people.

An effective malaria vaccine could prevent many cases and slow the problem of evolving drug resistance. But the complex life cycle of the malaria parasite makes it a difficult vaccine target [see "Molecular Approaches to Malaria Vaccines," by G. Nigel Godson; SCIENTIFIC AMERICAN, May 1985]. A bite from a mosquito introduces sporozoites, the first form of the parasite, into a human host. Within minutes, the sporozoites burrow into liver cells and start multiplying. In other forms, they destroy red blood cells and pass on the infection.

With each metamorphosis, Hoffman notes, the immunologic features of the parasite change, so vaccines that work against one form may not work against others. That subterfuge seems to help malaria defeat the immune system: although people who have been repeatedly infected with malaria over decades seem to build up a partial resistance to it, no one ever becomes naturally immune. "We have to do better than Mother Nature," Hoffman says.

More than two decades ago researchers demonstrated that they could prevent malaria in rodents and humans by injecting them with radiation-weakened sporozoites. Producing sporozoites in the laboratory is difficult, however. "We couldn't even begin to dream of growing enough to protect people in Africa," Hoffman says. Instead he and other vaccine researchers have in effect been trying to "rip the parasite apart to determine the targets and mechanisms of the irradiated sporozoite-induced immunity," Hoffman comments.

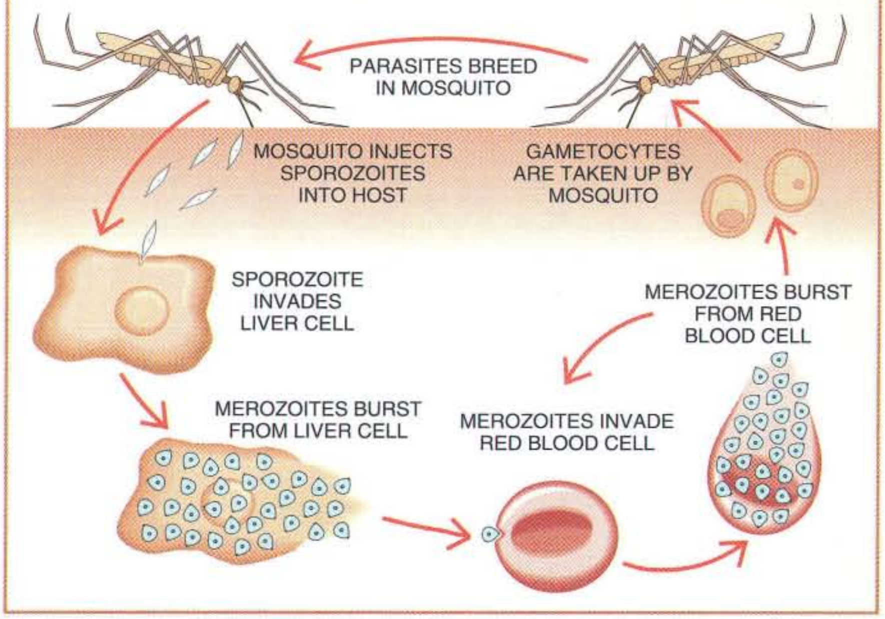
Much of the work on possible vaccines has concentrated on the circumsporozoite (CS) protein, a highly immunogenic feature of sporozoites. But according to Victor Nussenzweig of the New York University Medical Center, who has worked on several experimental vaccines, the CS protein–based products "did not produce the levels of antibodies that we expected to see. In animal models they worked beautifully, but in humans the levels were much, much lower." A second problem was that the vaccines were not designed to mobilize killer (cytotoxic) *T* lymphocytes, white blood cells that are a powerful component of the immune system. Active killer *T* cells are crucial to malaria immunity.

Hoffman's group was more successful because it did not place all its disease-fighting eggs in one immunologic basket. The team's vaccine presents two protein targets: the CS protein and a more recently discovered one called sporozoite surface protein 2 (SSP2). The group inserted the genes for those proteins into tumor cells and injected the modified cells into mice.

Many of the mice exposed to either protein in this way were protected against infections by sporozoites. Mice exposed to both proteins, however, were "100 percent" protected, Hoffman reports—even when they were challenged with extraordinarily large numbers of parasites. He believes that the vaccine stimulates *T* cells to attack the parasites in the liver. Antibodies produced against the CS and SSP2 proteins may also be killing the sporozoites.

Encouraging as the results are, the obstacles in adapting the mouse vaccine mean that many years—Hoffman refuses even to speculate on how many—will probably be needed before an analogous human vaccine is ready. SSP2 is a protein on a mouse malaria parasite; the equivalent protein on human parasites must still be found. Because injecting tumor cells into people is unacceptable, other effective ways of presenting the proteins must be developed. "So we have our work cut out for us," Hoffman says.

**The Life Cycle of the Malaria Parasite**


PARASITES BREED IN MOSQUITO
MOSQUITO INJECTS SPOROZOITES INTO HOST
GAMETOCYTES ARE TAKEN UP BY MOSQUITO
SPOROZOITE INVADES LIVER CELL
MEROZOITES BURST FROM RED BLOOD CELL
MEROZOITES BURST FROM LIVER CELL
MEROZOITES INVADE RED BLOOD CELL

Privately, some malaria researchers have expressed doubts that Hoffman's success in mice will do much to advance the development of a human vaccine. Nussenzweig has greater confidence: "My hunch is that it will be applicable," he says.

By using both the CS and SSP2 proteins, "have we done what the irradiated sporozoite vaccine does? I doubt it," Hoffman remarks. "I think there are more targets and more mechanisms." An ideal vaccine, in his opinion, would include proteins from every stage of malaria infection. He emphasizes that "none of us in this field believes a vaccine by itself will be able to control malaria": drug treatments and mosquito control measures will always be needed. *—John Rennie*

---

## Birds of a Fever

### *A lethal malaria may have an avian origin*

While vaccine researchers have been working toward eradicating *Plasmodium falciparum* malaria, the deadliest strain, one group of molecular biologists has been figuring out where it came from. Their conclusion: people probably caught the parasite from birds as an indirect consequence of agriculture. If so, the disease is less than 10,000 years old.

Parasitologists have noticed that *P. falciparum* is an oddball among the four species of parasites that cause human malarias. The organism is extraordinarily virulent: it accounts for more than 40 percent of all malaria cases and more than 95 percent of all malaria deaths. "For a parasite, killing the host usually isn't such a good idea," explains Thomas F. McCutchan of the National Institute of Allergy and Infectious Diseases (NIAID).

That observation suggests that *P. falciparum* has not yet fully adapted to its human host. Another difference is that the gametocytes, or sexual forms, of the other human malarias are round, whereas the gametocyte of *P. falciparum* has a sickle shape like those of avian malarias.

To test the relatedness of *P. falciparum* and other malaria parasites, McCutchan's group compared the nucleic acid sequences of RNA molecules found in the parasites' ribosomes. The number of differences in the RNA subunits offers clues to how the species may have branched off from one another.

As McCutchan and his colleagues recently described in the *Proceedings of the National Academy of Sciences,* the RNA comparisons verified their suspicions: *P. falciparum* is much more similar to avian malarias than to the other human malarias. In contrast, *P. vivax,* a less virulent human malaria, is clearly related to the simian disease and could have coevolved with humans.

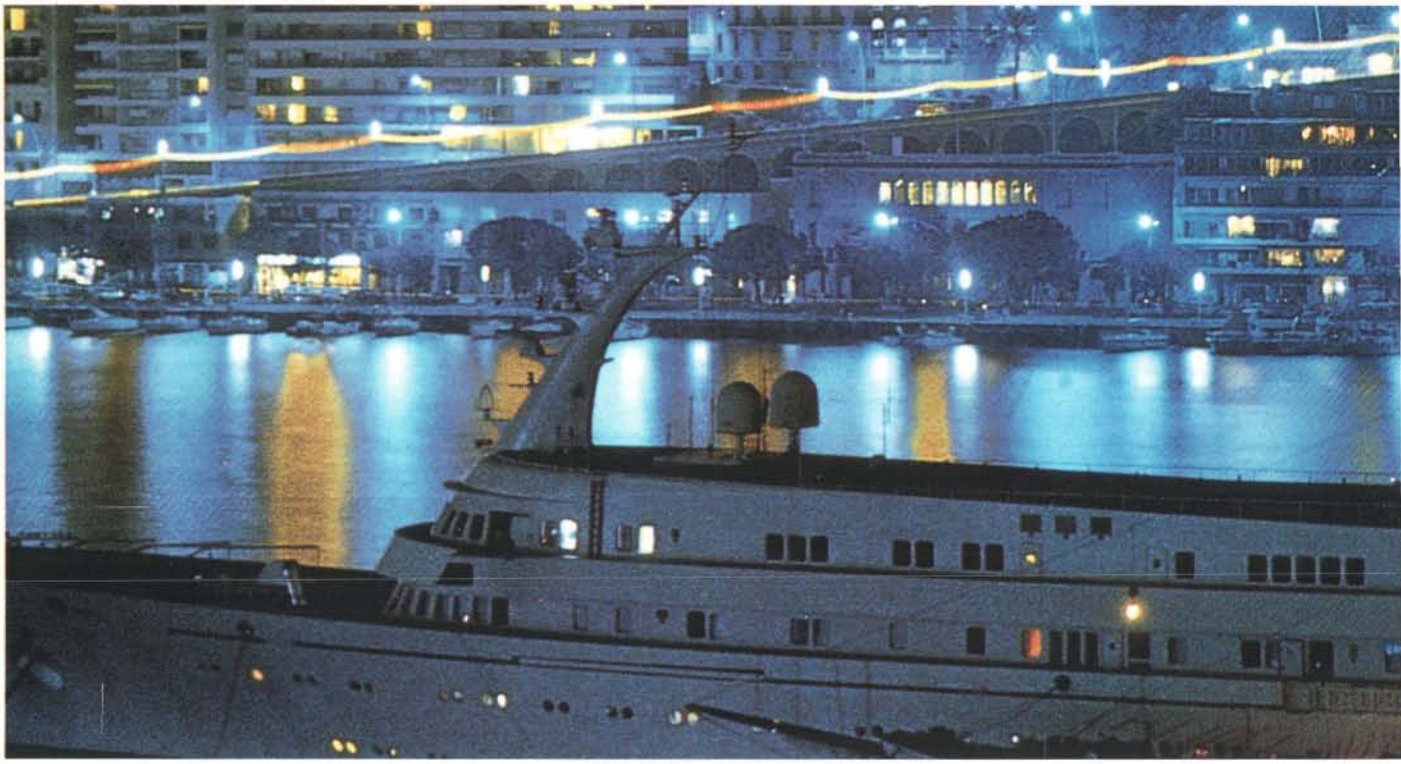
According to McCutchan, the best explanation is that *P. falciparum* malaria is the result of a lateral transfer from birds. Long ago a mosquito that had bitten a malaria-infected bird may then have bitten and infected a human being. Most parasites would not have survived; unfortunately, the ancestor of *P. falciparum* was an exception.

Such a transfer could have taken place at any time, but McCutchan argues that the rise of agriculture was critical because agrarian societies would be stationary long enough for the disease to take hold. "The disease life cycle couldn't really sustain itself with hunter-gatherers, because they would just wander out of the area," he says.

McCutchan also points to research by Mario Coluzzi of the Institute of Parasitology of the University of Rome, which shows that *Anopheles gambiae,* the mosquito primarily responsible for transmitting *P. falciparum* malaria, thrives almost exclusively where people live or work. Small, stagnant pools on cleared lands are ideal breeding areas for the mosquito.

It is also conceivable, McCutchan says, that the domestication of fowl as farm animals played a role in bringing infected birds and people together and increasing the opportunities for lateral transfer. Further studies may suggest which avian host might have harbored *P. falciparum's* nearest ancestor. *—John Rennie*

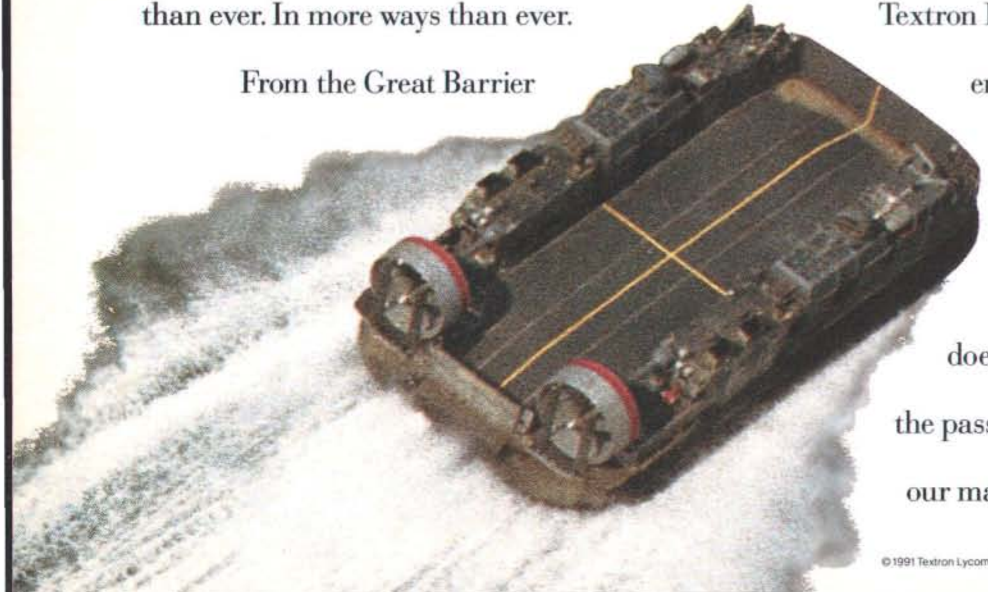
# From The Halls To The Shores



Textron Lycoming is hitting more beaches than ever. In more ways than ever.

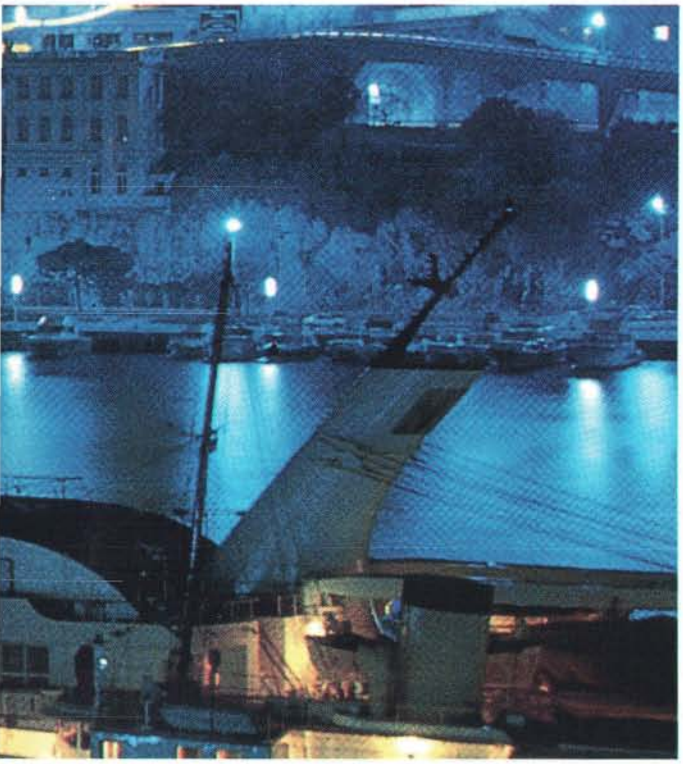
From the Great Barrier Reef to the Strait of Gibraltar to the Isles of Scilly. Textron Lycoming's TF Series marine turbine engines power all kinds of ships—from assault craft to megayachts—all over the world.

Camouflaged Marines or tuxedoed millionaires. The sea doesn't care who the passengers are. Neither do our marine turbines. They'll



©1991 Textron Lycoming

# Of Montezuma Of Monaco.



carry them all quickly and quietly. With a minimum of maintenance, and no vibration or smoke.

In the U.S. Navy's Landing Craft Air Cushion (LCAC), designed and built by Textron Marine Systems, the power of our turbines lifts and propels over 60 tons at speeds in excess of 40 knots. In the megayacht "Fortuna," it keeps the king of Spain and his guests moving comfortably through the Mediterranean. In the speedboat "Gentry Eagle," it provided the boost to break a world record in the rugged waters of the North Atlantic. And in fast, profitable commercial ferries in many ports, Textron turbines have provided the dependable power that has delivered thousands of passengers safe and on time.

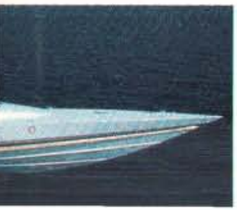
The Textron Lycoming TF Series marine turbine engine can do a lot of things. In a lot of places. What can it do for you?

There's only one way to find out. Call our Advanced Applications Department at (203) 385-2255. If you've got an idea that needs momentum —on sea, land, or air—call Textron Lycoming, U.S.A.

## We're On The Move.

**TEXTRON Lycoming**

Textron Lycoming/Subsidiary of Textron Inc.



# Molecular Trickster

*Antisense RNA pulls a fast one on the leukemia virus*

Like all viruses, lethal retroviruses that cause diseases such as AIDS and leukemia reproduce by tricking cells into making myriad copies of themselves. But researchers at Ohio University have played a molecular prank of their own. They altered the cells of mice to produce sequences of RNA that fool a leukemia virus into making duplicates that cannot infect other cells.

Although the technique cannot be applied to humans, the study is the first published work to show that so-called antisense RNA can block retroviral infection not just in test tubes but in living organisms. "The antisense totally shuts off viral replication," says Thomas E. Wagner, scientific director of the Edison Animal Biotechnology Center at Ohio University. "It's not a question of less—there was absolutely none."

A relative of the AIDS virus (HIV), the leukemia retrovirus carries its genetic information in the form of RNA, rather than as the double-strand DNA employed by other organisms. Once the pathogen has invaded a cell, it uses an enzyme called reverse transcriptase to synthesize DNA that is then incorporated into the nucleus of the host cell. (This process of reverse transcription from RNA into DNA gave rise to the term "retrovirus.")

After taking over the cell's replicative machinery, the retroviral DNA directs production of new viral particles. A key part of this process is a short segment of RNA called the packaging sequence. It enables two identical copies of the viral genome to wrap around core proteins; the pairs become the contents of new viral shells that exit the cell in search of other victims.

The Ohio University scientists used antisense RNA to block the packaging process. The trick was to alter the DNA in the mouse cells so that it produced RNA that was an exact mirror image of the viral RNA sequence that controls packaging of the new viral genome. The antisense RNA sticks like Velcro to the viral RNA, preventing it from binding with the packaging protein. New viral particles still emerge from the invaded cell, but without the genetic information necessary to infect other cells.

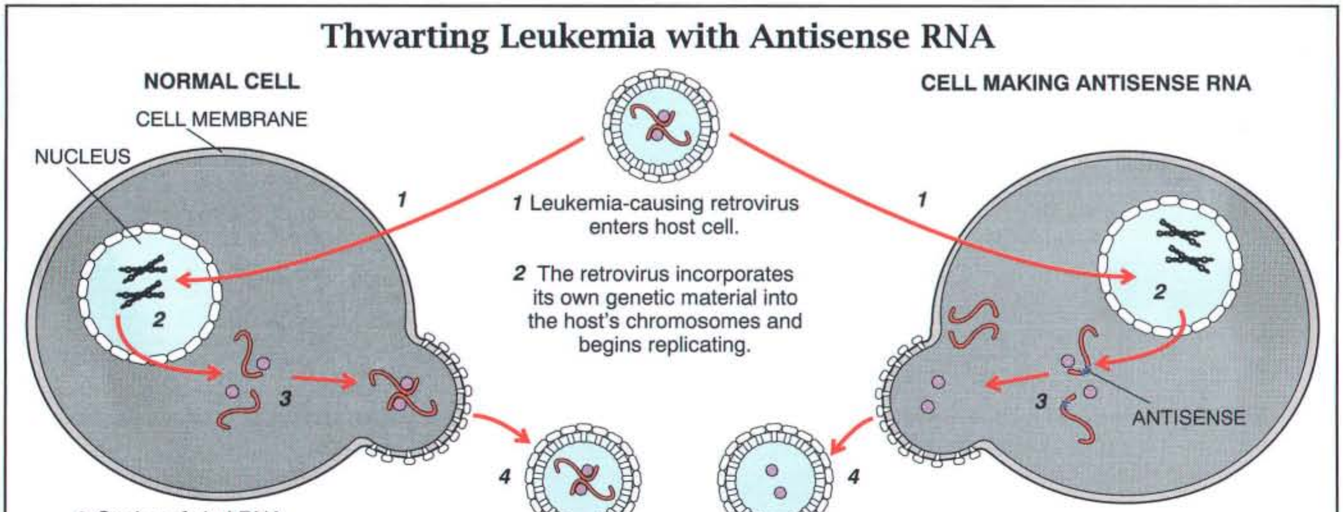
"The empty viral shells do nothing and quickly degenerate," explains Lei Han, who conducted the experiment in Wagner's laboratory as part of her doctoral studies. To enable cells to replicate the antisense RNA, Han borrowed retroviral promoter and terminator sequences—molecular tools that the virus uses in copying its genome—and used these to flank the synthetic RNA. The pieces caused the strand to coil into a circular vector, or ferrying unit, that was microinjected into fertilized mouse eggs. Embryos that accepted the material developed into mice with cells that produced copies of the antisense sequences.

In a paper published in the May 15 issue of the *Proceedings of the National Academy of Sciences,* the investigators reported that when the transgenic mice were exposed to murine leukemia virus (a type that cannot infect humans), none acquired the disease. Some 31 percent of a control group did, and many others were showing signs of infection when the experiment was halted. "We had to stop when we did, because the controls were all dying. You can't analyze a dead mouse," Wagner explains.

This type of prenatal gene therapy will not be considered for humans, but it does point to the potential benefits of a similar approach. "If you take a mature animal and can engineer its hematopoietic [blood-forming] system, it will give rise to an entire army of resistant cells that will repopulate the body," declares Nava Sarver, chief of targeted drug discovery for the division of AIDS at the National Institute of Allergy and Infectious Diseases. HIV is a potentially good target, she notes, because the retrovirus attacks only *T* cells.

One way to protect *T* cells would be to introduce antisense into stem cells, the bone marrow cells that differentiate into immune and red blood cells. Stem cells have not yet been retrieved from humans, but scientists expect it will be possible within a few years.

"Now that people know that getting antisense RNA into the appropriate cells will work, they will be more motivated to develop delivery techniques that are clinically useful," says W. French Anderson, chief of molecular hematology at the National Heart, Lung and Blood Institute, who headed the first medical group to perform human gene therapy last September. Han, with her freshly minted Ph.D., has already gone to join his team. —*Deborah Erickson*



**Thwarting Leukemia with Antisense RNA**

*1* Leukemia-causing retrovirus enters host cell.

*2* The retrovirus incorporates its own genetic material into the host's chromosomes and begins replicating.

*3* Copies of viral RNA wrap around core proteins.

*4* New viral particle buds forth to infect another cell.

*4* Because it lacks critical genetic material, the new particle is incapable of infecting another cell.

*3* Pieces of antisense RNA stick to the packaging sequence in copied viral RNA, preventing binding to core proteins.

## Quantum Magnets

*Will quantum effects wreak havoc on magnetic data storage?*

If your magnetic computer disk suddenly erases itself, don't curse the equipment. Blame it on quantum mechanics. The theory claims that when a tiny magnet is cooled to absolute zero, the north pole and the south pole can effortlessly switch position. To a lesser extent, this effect, known as magnetic quantum tunneling, should affect relatively large, warm magnets, such as those that make up a data disk.

To be sure, you are much more likely to destroy the data by jolting the disk with static electricity than by having magnetic domains suddenly change polarity. But at least some physicists are beginning to wonder what role quantum mechanics might play in small magnets, particularly in the next generation of dense data storage systems. From 1950 to 1990, the number of atoms required to store a single bit of information dropped from 100 billion billion to one billion. If progress continues at that rate, magnetic storage systems should pass the mark of 100,000 atoms per bit before the year 2000.

David D. Awschalom and his colleagues at the IBM Thomas J. Watson Research Center have pioneered techniques to study magnets composed of 100,000 atoms at temperatures approaching absolute zero. They can measure a magnetic field with a million times more sensitivity than can conventional techniques, such as nuclear magnetic resonance. And they believe they may have observed evidence of tunneling and other effects of quantum mechanics in these magnetic systems.

The work raises fundamental issues in both physics and engineering. Classical physics governs the behavior of a magnet composed of a billion atoms at room temperature; quantum mechanics sets the rules for a one-atom magnet at absolute zero. So it is still anyone's guess as to how the two theories should cope with magnets that are not so small and not so cold.

The key to Awschalom's success is the ability to integrate tiny magnets into a device called a microsusceptometer. The device consists of two square loops of wire, each 20 microns on a side. These so-called pickup loops are connected so that when an electric current flows clockwise through one loop, it will move counterclockwise through the other. Tiny magnets are deposited within one of the pickup loops. The magnets give rise to a very weak current in the loop. The current is amplified using a superconducting quantum interference device, commonly known as a SQUID.

Each pickup loop is surrounded by a larger loop. When a current flows through these two larger loops, they generate two magnetic fields that have the same strength and orientation. The fields will then induce the same amount of current to flow through each of the pickup loops. But because the loops are counterwound, the currents will circulate in the opposite direction and cancel. Only the current induced by the tiny magnets flows through the pickup loops. Hence, the pickup loops do not sense the applied fields but rather the influence of the applied fields on the tiny magnets.

To deposit the magnets within one of the pickup loops, the IBM group used a scanning tunneling microscope. The chamber of the microscope was filled with a vapor of iron pentacarbonyl, Fe(CO). When a voltage was applied between the tip of the microscope and the sample, the vapor would break down, placing a droplet of iron and carbon on the sample. The droplet consists of 100,000 tiny magnets, or spins.

The IBM group placed 100 or so droplets within the pickup loops. (The advantage of working with 100 droplets rather than just one is that their combined magnetic field is more easily detected.) The droplets were spaced far enough apart so that the magnetic field of one did not influence the behavior of droplets nearby. When the 100,000 spins in each droplet are cooled to 20 millikelvins, two hundredths of a degree above absolute zero, each spin locks to the one next to it. The entire droplet acts as one gigantic spin whose north pole points either up or down.

Classical mechanics predicts that the spins cannot change their orientation without receiving some energy. This energy barrier should prevent spins from flipping around. Quantum mechanics states that there is a chance that the spins will be able to break through the barrier. (This phenomenon is analogous to the ability of an electron to tunnel through an energy barrier.)

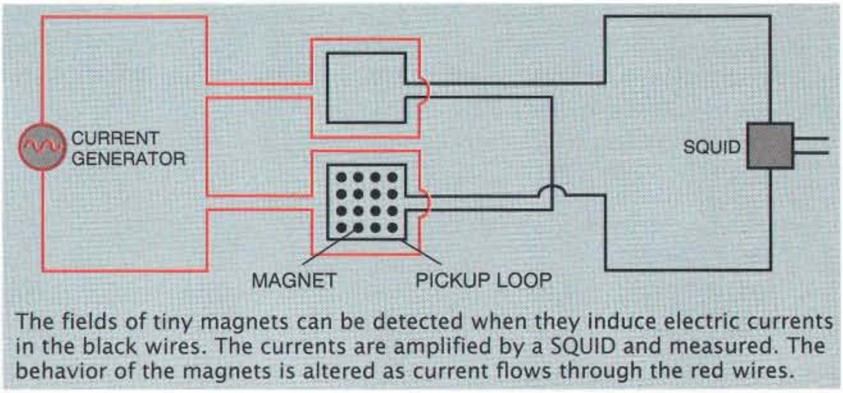
To test the theories, the IBM workers first measure the magnetic properties of the droplets in the absence of a magnetic field. If quantum mechanics applies in this case, all 100,000 spins in each droplet should spontaneously flip up and down together at a certain frequency. The investigators then introduce a magnetic field whose polarity switches up and down over time. If the frequency of the applied field matches the natural frequency of the magnets, then the flipping will be enhanced.

This kind of resonance effect was just what the IBM team observed. This experiment and several others seemed to make a strong case for quantum magnetic tunneling. "We were jubilant," Awschalom cheers. "We thought, 'Ah, we've seen it,' the first real observation of quantum mechanics playing a role at the macroscopic level" in magnetic systems.

But as Awschalom readily admits, theory and experiment do not resonate. The theory suggests a way to derive the natural frequency of tunneling from such quantities as the size of the magnets, their temperature and the strength of the applied fields. The calculated frequency is a million times greater than what the IBM group measured. "We can't get all the numbers to jibe," Awschalom explains. "What we've seen is either a completely new phenomenon, which is possible, or the theory is incomplete, which is also possible."

In either case, Awschalom and his collaborators face the difficult task of explaining the odd behavior they have found in tiny magnets. If they can blame it on quantum mechanics, who knows what's in store, or can be stored, in magnets. *—Russell Ruthen*

### A Device for Observing Quantum Magnets



The fields of tiny magnets can be detected when they induce electric currents in the black wires. The currents are amplified by a SQUID and measured. The behavior of the magnets is altered as current flows through the red wires.

## Math Exorcise

*A new algorithm lifts the curse of dimensionality*

Few demons plague science, statistics and homework like solving for *X*. Although the task can be trivial, integrating a function that consists of multiple variables frequently proves intractable. One class of such multivariate integrals remained completely impervious to mathematical attack for more than two decades.

Then, Henryk Woźniakowski, a mathematician and computer scientist with a joint appointment at Columbia University and the University of Warsaw, derived an algorithm that breaks what mathematicians call the "curse of dimensionality," at least for integration. The result, which is attracting interest from physicists, chemists, biologists, engineers and even Wall Street rocket scientists, promises to make the computation of complex functions faster and more efficient.

The curse refers to the increase in the difficulty of a problem as the number of variables, or dimensions, increases. "If you want four-place accuracy, then every time you increase the dimension by one, you make the problem 10,000 times as hard," says Joseph F. Traub, a computer scientist at Columbia well regarded for his research on algorithms and computational complexity. Complicated systems such as chemical reactions typically involve thousands of dimensions. An exact solution becomes impossible to obtain.

Instead investigators approximate the solution by using computers to calculate the integral numerically. The most familiar way to get an answer is with a technique known as the Monte Carlo method: to approximate the solution of an integral, evaluate the integrand at randomly selected points and then extrapolate. The number of points selected represents the "cost" of the computation. The more points chosen, the closer the approximate value will be to the true value. Unfortunately, using more points also lengthens the computational time.

Still, some functions are thought to be so complex that the cost of approximation would be huge even for a moderately large error (represented by the Greek epsilon, $\varepsilon$). Such complex problems, however, would be easier to solve in an "average case" setting. Whereas the Monte Carlo method guarantees the answer within some $\varepsilon$, an average-case algorithm offers solutions with small expected, or average, errors. (In essence, Monte Carlo is more of a worst-case scenario; the answer falls within the maximum error.)

Although obtaining answers within some average error may seem less desirable than getting them within some maximum error, average-case answers are not necessarily any "worse." Mathematicians had proved that an average-case algorithm guarantees an error at least as good as the maximum error, depending on which sample points are used to approximate the solution. Unlike Monte Carlo, which uses randomly selected points, the average-case algorithm uses deterministic, or nonrandom, points. Finding those particular sample points, however, has proved elusive. All previous attempts failed because of dimensionality.

Woźniakowski's task, then, was to find an efficient sampling algorithm that approximates the integral at minimal average cost—without falling under the curse of dimensionality. The key breakthrough came when Woźniakowski realized that the average-case problem is mathematically equivalent to a problem in number theory known as discrepancy.

Luckily, Klaus F. Roth, a Fields medalist from the Imperial College in London, solved the discrepancy problem in 1954. Woźniakowski found that "it's possible to translate Roth's result to the problem of multivariate integration." The translation makes it possible to select the sample points that make the average error small.

The algorithm is much faster than the Monte Carlo method. The number of points needed to guarantee an average error less than some specified error $\varepsilon$ is roughly proportional to $1/\varepsilon$. In contrast, the number of points needed for the Monte Carlo method is proportional to $1/\varepsilon^2$. (This result also seems to imply that for the same number of sampling points, Woźniakowski's method is more accurate than Monte Carlo. But Traub cautions that "it's a different error criterion, so one has to be careful not to compare apples with oranges.")

Woźniakowski's work, which appeared in the *Bulletin of the American Mathematical Society* in January, "is a beautiful mathematical result that has settled an open question in complexity theory," Traub remarks. Indeed, he notes that the paper has stimulated interest in many different disciplines. John R. Birge, who studies optimization problems at the University of Michigan, points out that the scheme could help financiers "design a portfolio to optimize the expected return." The sample points would be the interest rates or prices of stock and commodities.

But for Woźniakowski, a theoretician, "it was the challenge of solving a long-standing problem. That's what I like." *—Philip Yam*

## Live from Off-Center

*Astronomers follow the energetic trail of the Great Annihilator*

No, it is not the new summer movie starring Arnold Schwarzenegger. The Great Annihilator is the deliberately colorful name that Marvin Leventhal of AT&T Bell Laboratories has bestowed on a puzzling and highly variable source of powerful gamma radiation located near (but not at) the center of the Milky Way.

A series of balloon-borne and satellite observations has revealed that the source sporadically radiates strongly at an energy of 511,000 electron volts. That number is precisely the amount of energy released when an electron encounters and destroys its antimatter twin, suggesting that the observed radiation results from such annihilation. The total amount of radiation from the source is enormous: in a single day last October it equaled 50,000 times the sun's total luminosity.

Gamma-ray telescopes traditionally have poor resolution, because gamma rays cannot be focused in the ways that visible light can. Nevertheless, observations over the past decade clearly demonstrated that something strange was going on near the galactic center. In 1980 the National Aeronautics and Space Administration's *HEAO-3* satellite revealed that over the course of a few months, the total brightness of the gamma radiation fell by two thirds. The *Gamma Ray Imaging Spectrometer,* a balloon-borne instrument developed at the NASA Goddard Space Flight Center, has examined the galactic center at very precise energies and has revealed that the brightness of annihilation radiation from the central region is particularly variable.

Astronomers had proposed several possible gamma-ray sources near the galactic center as the source of the Annihilator's radiation. The Soviet *GRANAT* satellite, which offers resolution superior to that of previous gamma-ray telescopes, seems to have ended the search. *GRANAT* revealed that the blast of annihilation radiation observed on October 14 came from the abstrusely named object 1E1740.7-2942, which earlier studies had identified as a source of less energetic X rays.

When *GRANAT* examined this object two days later, the annihilation component had vanished. That finding puts a distinct limit on the size of the Great Annihilator, because an object's brightness cannot vary faster than the time it takes light to cross its diameter. Therefore, the Annihilator must be no larger than two light-days (about four times the diameter of Pluto's orbit about the sun), and it could be far smaller.

When astronomers try to imagine a compact but very potent source of radiation, they inevitably think "black hole." Indeed, Leventhal and John Bally, also at Bell Labs, speculate that the Annihilator is a solitary black hole of no more than 100 solar masses. The radiation comes not from the hole itself but from nearby gas and dust as it swirls into the hole, grows extremely hot and glows intensely. At times the material around the hole radiates so powerfully that some radiation spontaneously creates pairs of electrons and positrons. When these particles recombine, they create the observed 511,000-electron-volt signal.

Bally finds distinctive microwave emissions from the vicinity of the Annihilator that suggest it is surrounded by a cloud of carbon monoxide and other molecules. This cloud could be the source of fuel for the black hole.

In some ways, the latest findings raise more questions than they answer. Many astronomers suspect that a far more massive black hole lies at the exact center of the galaxy, and yet there is no sign of it at gamma-ray energies. And nobody knows what causes the brightness of the Great Annihilator to vary so wildly.

Examining the Annihilator at other wavelengths may help explain its fickle behavior and provide more evidence that it really is a black hole. Thomas A. Prince of the California Institute of Technology has detected radio emissions that he thinks also emanate from the Annihilator. The German-American satellite *ROSAT* will observe X-ray emissions from the object to see whether they correlate with its behavior at other wavelengths.

The quest to understand the Annihilator has also been joined by NASA's sophisticated *Gamma Ray Observatory,* or *GRO,* which was launched this past April. *GRO* is the second of four planned "Great Observatories," which will examine the sky at infrared, visible, X-ray and gamma-ray wavelengths. Unlike the *Hubble Space Telescope,* the first Great Observatory to be launched, *GRO* went up accompanied by relatively little public fanfare. Also unlike *Hubble, GRO*'s performance "is beyond design specifications," says *GRO* project scientist Donald A. Kniffen of the NASA Goddard Space Flight Center.

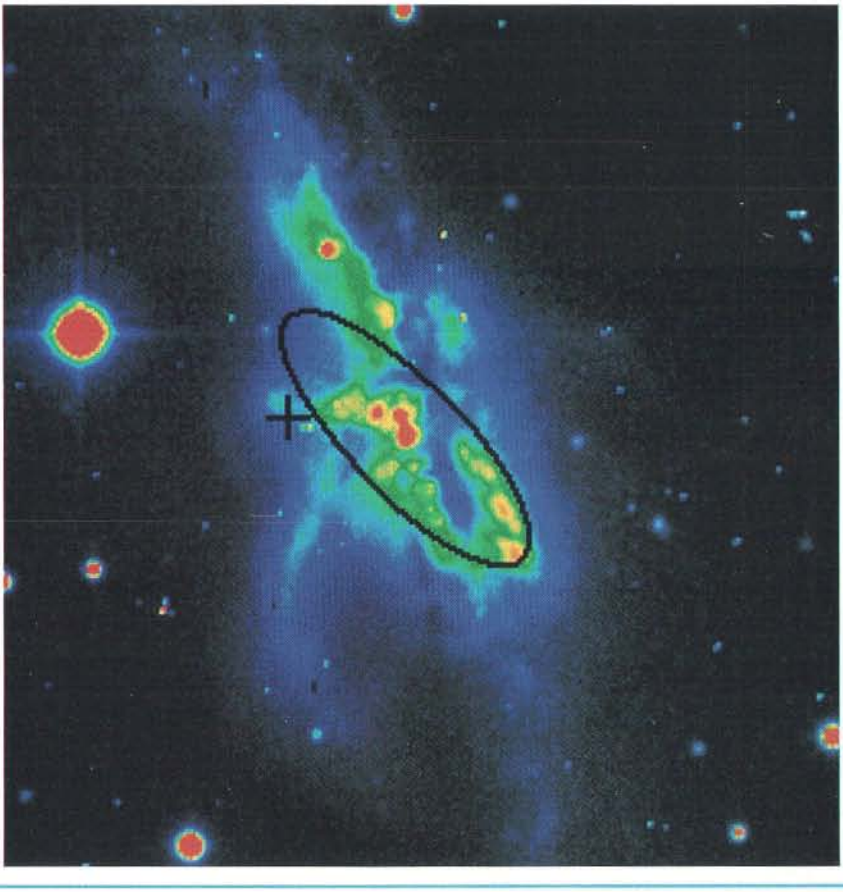
One of *GRO*'s four instruments, the *Oriented Scintillation Spectrometer Experiment* (*OSSE*), will be particularly useful for monitoring the ups and downs of the Annihilator. *OSSE* will watch the Annihilator for very rapid fluctuations in brightness and will make sensitive measurements of the energy of the radiation it emits. Among *GRO*'s first targets will be an object known as Cygnus X-1, another likely black hole candidate whose behavior is in some ways similar to that of the Annihilator.

For the moment, the Annihilator seems to be in a quiet phase. But if its past behavior is any indication, the words of Schwarzenegger should be appropriate: "I'll be back." Astronomers will be waiting. *—Corey S. Powell*

## Biggest Black Hole in the Universe?

Some 300 million light-years from the earth, next to a spiral galaxy with the prosaic name NGC 6240, lurks something very dark and very heavy. Examining the galaxy with a telescope in Hawaii, astronomers noticed a whirlwind of stars and gas just beyond the galaxy's edge. The astronomers calculated that at the center of the cyclone lies something as massive as the entire Milky Way—which contains 100 billion stars—but at least 10,000 smaller in volume and emitting no detectable light. In this image, an ellipse shows the outline of NGC 6240, and a cross marks the location of the massive object.

The astronomers, who come from Rice University, the University of Maryland and the University of Hawaii, think the object is probably a black hole. If so, it would be the biggest ever hypothetically observed (all black holes are still hypothetical). But the investigators, who published their findings in the *Astrophysical Journal,* note that the object could also consist of a dense swarm of neutron stars or brown dwarfs (which are somewhat less hypothetical). *—John Horgan*



# PROFILE: THE DALAI LAMA

## A Subtle Mind Contemplates Science

The monk in maroon robes and sunglasses decides not to talk about watches, one of his favorite subjects. He sets aside a prepared speech in which he relates his awe for the U.S. to his awe for the beauty of a gold watch that President Franklin D. Roosevelt once sent him as a gift, and, with palms joined together, he urges a roomful of congressional leaders to work toward peace and human rights. The audience in the Capitol's Rotunda is quiet, reverent, listening to the Dalai Lama, who every now and then—it almost seems as if he can't help it—glances at his watch.

The Dalai Lama's fascination with technology is by no means limited to timepieces. This spring, on his most recent trip to the U.S., he met not only with politicians—including President George Bush—and Buddhist followers but also with prominent scientists. The exiled Tibetan leader, who won the 1989 Nobel Peace Prize for his campaign for world peace and for Tibetan freedom, has repeatedly sought out astronomers as well as physicians, psychologists and neuroscientists.

On this trip the Dalai Lama participated in a conference on neuroscience and psychology sponsored by Harvard Medical School, New England Deaconess Hospital and Tibet House. He also visited Cornell University, stopping by the computer laboratory of Donald P. Greenberg, where a Tibetan monk is working on a software program to render a mandala, a religious symbol, in three dimensions. When he saw the mandala, a research associate there says, the Dalai Lama burst into giggles of delight, something he does even more often than looking at his watch.

The West can learn from Tibetans, too, the Dalai Lama says, particularly in the field of neuroscience. To that end, in 1987 he and Francisco J. Varela, a neuroscientist at Centre National de la Recherche Scientifique in France, began a series of meetings called Mind and Life, where researchers have presented their work to the Dalai Lama. Several of these scientists are now planning studies that could use Tibetan meditative practices—developed over 2,500 years—as tools for conducting research. During meditation, a monk slows down physical processes, concentrating on what is called subtle mind, a term used to describe mind stripped of conceptualizations and sensory perceptions, says B. Alan Wallace, a graduate student at Stanford University and former Tibetan monk, who translates for and helps to organize the conferences.

The Dalai Lama also believes Buddhist philosophy is relevant to environmental studies because it acknowledges the interdependence of all living things. On the day after his address to members of Congress, the Dalai Lama spoke in a small, plush room at the National Museum of American History to another rapt audience, which included Senator Al Gore of Tennessee and David R. Brower, former executive director of the Sierra Club and founder of Earth Island Institute. Cracking jokes every few minutes, the Dalai Lama called for the gentle treatment of all creatures—although he admitted to having committed a few understandable sins against mosquitoes.

Born as Lhamo Dhondrub in 1935 among the snow-covered peaks at the roof of the world, the Dalai Lama would seem to be an unlikely student of Western science. He was recognized by the Tibetan people as a god-king at the age of two and was taken from his small village to a drafty 1,000-room palace in the Tibetan capital to learn the affairs of state and Buddhist philosophy.

As the ruler apparent, he had access to the few gadgets that made their way over the Himalayas. He remembers being enthralled by a Meccano set, a telescope, an electric generator and three automobiles—all quite rare in Tibet. Although he laughs at the idea of being considered an expert, the Dalai Lama explains that he used to repair the palace's manual movie projector. And he modestly admits to a talent for repairing watches (although President Roosevelt's golden gift proved too intricate and had to be fixed in Switzerland).

The Dalai Lama's training was largely philosophical. By the time he assumed political responsibilities in 1950 at the age of 15, his studies of Buddhist philosophy were well under way—and 10 years later he earned a degree equivalent to a doctorate. But the political events that forced him out of the remote and mysterious land of Tibet provided him greater contact with science. By 1959, after Tibet had suffered a decade of increasingly violent attacks by the Chinese, the Dalai Lama and some 87,000 followers fled. Since then, he has lived in Dharamsala, a town in the Himalayan foothills of northern India.

The Dalai Lama, whose every visit to the U.S. is accompanied by protests from the People's Republic of China, concedes there is a diplomatic dimension to his scientific exchanges. Tibetan civilization was nearly destroyed during the Cultural Revolution: more than 6,000 monasteries were razed, scores of ancient texts were burned and thousands of Tibetans, including monks and nuns, were killed. If Tibetan practices prove valuable to scientists, the Dalai Lama believes, then the Chinese government may become convinced of the uniqueness and importance of Tibetan culture.

"So you see, I am not trying to antagonize the Chinese but to enlighten their ignorance," the Dalai Lama exclaims. "That's my strategy. A little bit of politics is all right," he adds, laughing uproariously during an interview in his Washington hotel room. Although his English is very good, the Dalai Lama often speaks with the help of a translator, Geshe Thupten Jinpa. Throughout the interview, however, he often interrupts—in perfect English—to clarify his points as they are being translated.

The Dalai Lama's first scientific collaboration, of sorts, came in 1979, when Herbert Benson, a professor of medicine at Harvard University, asked permission to study metabolic rates in meditating monks. Benson, who had also studied physiological changes during transcendental meditation, reported a 64 percent reduction in oxygen consumption during certain meditative states in the subjects. Benson and his colleagues also described a practice of Tibetan monks called *thumo:* generating sufficient body heat to survive freezing weather. *Thumo* was reported in the 1920s by Alexandra David-Neel, a Frenchwoman who made her way through Tibet disguised as a Tibetan pilgrim.

The Mind and Life meetings grew out of the Dalai Lama's developing interest in psychology and neuroscience. The first was held in India, followed two years later, in 1989, by a second in California. A third convened last November in Dharamsala. "There was a gulf in terms of belief, but it never stood in the way of good feeling and communication," recalls Larry R. Squire, a professor of psychiatry at the University of California at San Diego, who presented research on memory and amnesia at the 1989 meeting.

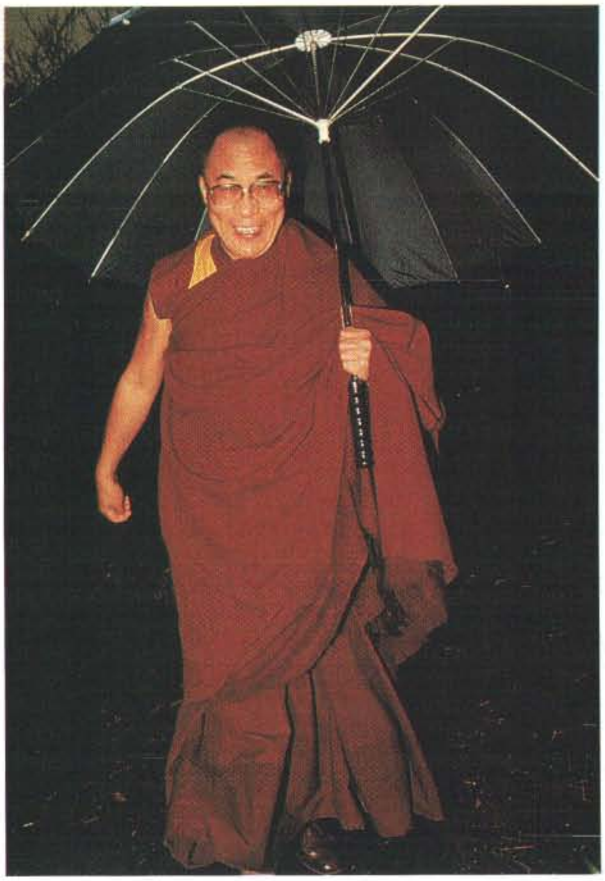
"It would be fair to say we did not make headway on what I would call the fundamental issues," including what mind is and what the brain does, Squire says, but he adds that Tibetan Buddhism "is another source of hypotheses that you could test. One doesn't just dismiss a system of belief that has been developed over thousands of years."

The Dalai Lama says that the Tibetan concept of mind is indeed a difficult one for some Westerners. To Tibetans, mind is not the brain. Instead certain states of mind are separable from the body. "On this particular point it is very difficult. Scientists are so far quite neutral," the Dalai Lama says good-naturedly, arranging the folds of his robe over his shoulder. In contrast, he adds, some scientists are attracted to other aspects of Tibetan thought: the absence of a divine god and of a permanent, unchanging soul.

In simple terms, Tibetan Buddhism holds that nothing exists intrinsically unto itself but rather that all phenomena—physical and mental—are dependently related events arising from causes, explains Wallace of Stanford. Consciousness, or mind, is regarded as a continuum, arising from previous consciousness. Tibetans apply this principle to birth and death, which gives rise to the belief in reincarnation, Wallace says.

Despite tenets such as reincarnation, the Dalai Lama says Tibetans and scientists have a lot in common. "The basic Buddhist attitude is that you should investigate and analyze," he describes, holding his left hand still while the right weaves over and under it, a gesture reminiscent of Tibetan debates in which monks accompany their arguments by clapping the right hand against the left. In Tibetan Buddhism, ideas—usually philosophical in nature—become clear or convincing through logic or through experience, "so the scientist's attitude and the Buddhist's attitude at that level are, I think, very similar," the Dalai Lama explains.

So far the Dalai Lama says he has learned nothing from the West that challenges Tibetan Buddhist doctrines. "On the contrary, we find scientific knowledge and findings very helpful," he observes. Some of these ideas come from the realms of quantum physics and cosmology. In Tibetan Buddhism, for example, the role of the observer cannot be separated from the observed: an idea central to quantum mechanics, notes Wallace, who studied physics at Amherst College. With regard to cosmology, an ancient Buddhist text depicts the universe as an oscillating one, describing expansion and collapse in a manner that recalls some versions of the big bang theory, according to Wallace.



*THE DALAI LAMA, the exiled Tibetan leader, says he is a scientist at heart. Photo: Steve Lehman, Visions.*

Some scientists who have met with the Dalai Lama feel cautious as well as intrigued. "The minute you say you like the Dalai Lama, your colleagues become worried about you," comments J. Allan Hobson, a psychiatrist at Harvard Medical School, who attended the second Mind and Life meeting. "I'm not implying that he wants to create groupies," Hobson adds, but "he may have mastered the art of unconscious proselytizing."

Hobson, whose own work centers on dreaming, was interested in the Tibetan view of dreaming as a second, lucid level of consciousness that some monks claim they can manipulate. But he said that there is "a strong need for the rigorous testing of some of those ideas." Hobson wants to study dream states in Tibetan monks but says so far he has not received clear guidance on how to proceed with such investigations. "It should be useful to talk about the study of consciousness. Now the question is: Where's the beef?"

The beef may be forthcoming. With the Dalai Lama's blessing, some scientists have put together a research network and have started to design studies to conduct in India, says Clifford Saron, a graduate student in neuroscience at the Albert Einstein College of Medicine in the Bronx. At the most recent Mind and Life session, Saron filled in for Richard J. Davidson, a professor of psychology at the University of Wisconsin at Madison, presenting research on emotions and brain laterality.

Saron, Varela and their colleagues are among the scientists who would like to conduct studies measuring the electrical activity in the brains of Tibetan monks to determine the effects of advanced meditation on brain function. Researchers also plan to examine why Tibetan torture victims reportedly do not exhibit posttraumatic stress, a psychological disorder characterized by nightmares and anxiety that has afflicted Vietnam veterans, among others. And since the immune system has been clearly linked in the West to the central nervous system and to emotional states, Tibetan practices could perhaps illuminate some cognitive mechanisms important to health, Saron says.

Such research may be difficult. It entails lugging heavy equipment to the other side of the globe, through India and up the winding road that encircles the lush, steep mountain on which Dharamsala is built. Highly sensitive electronic devices and computers will have to find their way into monasteries and into remote caves, where monks have remained in solitude for years. These two worlds have already come together for the Dalai Lama, however. On his wrist he wears a black digital watch next to his *treng-nga*, or Tibetan Buddhist rosary. —*Marguerite Holloway*

SCIENTIFIC
AMERICAN

# Opium, Cocaine and Marijuana in American History

*Over the past 200 years, Americans have twice accepted and then vehemently rejected drugs. Understanding these dramatic historical swings provides perspective on our current reaction to drug use*

by David F. Musto

Dramatic shifts in attitude have characterized America's relation to drugs. During the 19th century, certain mood-altering substances, such as opiates and cocaine, were often regarded as compounds helpful in everyday life. Gradually this perception of drugs changed. By the early 1900s, and until the 1940s, the country viewed these and some other psychoactive drugs as dangerous, addictive compounds that needed to be severely controlled. Today, after a resurgence of a tolerant attitude toward drugs during the 1960s and 1970s, we find ourselves, again, in a period of drug intolerance.

America's recurrent enthusiasm for recreational drugs and subsequent campaigns for abstinence present a problem to policymakers and to the public. Since the peaks of these episodes are about a lifetime apart, citizens rarely have an accurate or even a vivid recollection of the last wave of cocaine or opiate use.

Phases of intolerance have been fueled by such fear and anger that the record of times favorable toward drug taking has been either erased from public memory or so distorted that it becomes useless as a point of reference for policy formation. During each attack on drug taking, total denigration of the preceding, contrary mood has seemed necessary for public welfare. Although such vigorous rejection may have value in further reducing demand, the long-term effect is to destroy a realistic perception of the past and of the conflicting attitudes toward mood-altering substances that have characterized our national history.

The absence of knowledge concerning our earlier and formative encounters with drugs unnecessarily impedes the already difficult task of establishing a workable and sustainable drug policy. An examination of the period of drug use that peaked around 1900 and the decline that followed it may enable us to approach the current drug problem with more confidence and reduce the likelihood that we will repeat past errors.

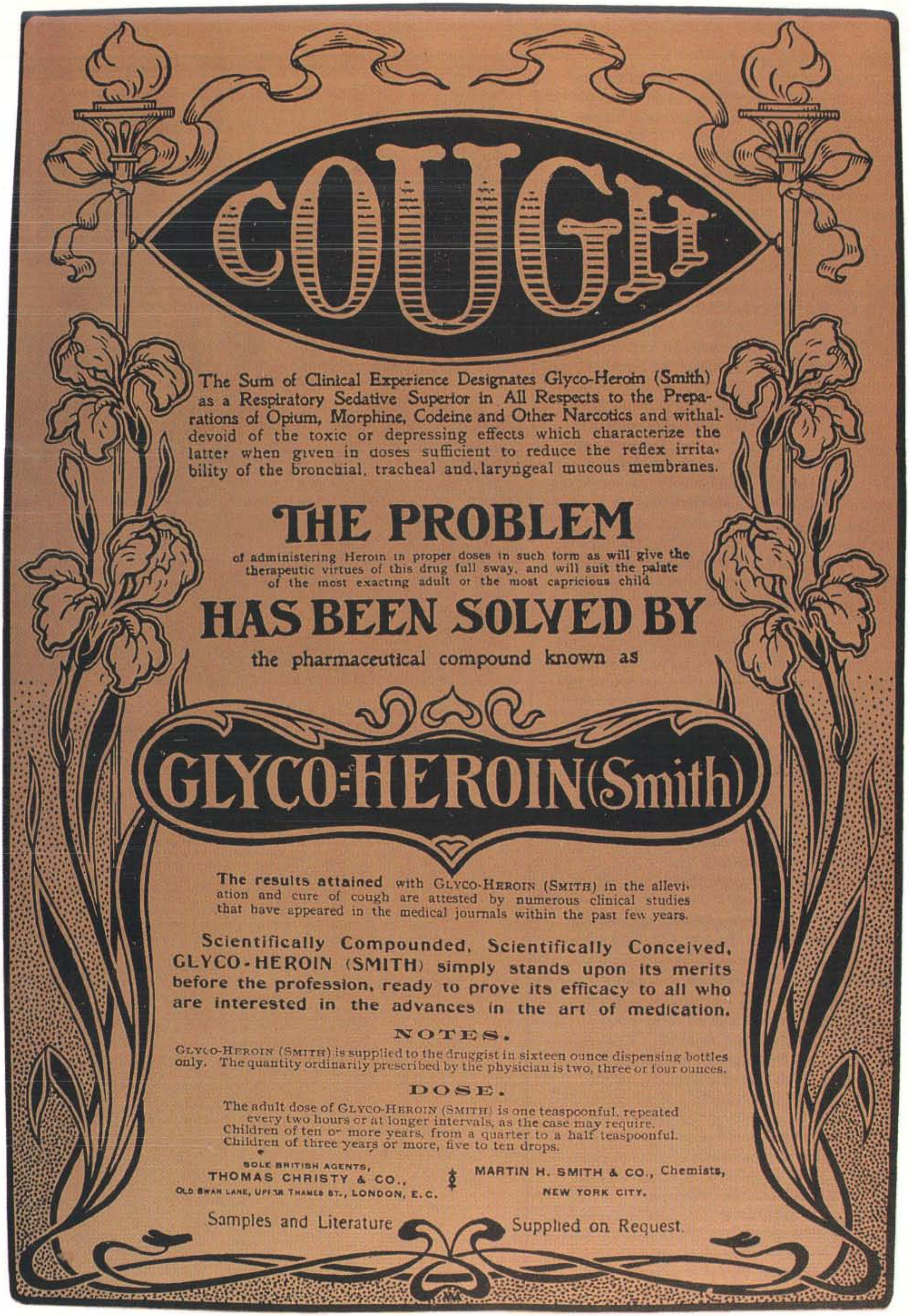
Until the 19th century, drugs had been used for millennia in their natural form. Cocaine and morphine, for example, were available only in coca leaves or poppy plants that were chewed, dissolved in alcoholic beverages or taken in some way that diluted the impact of the active agent. The advent of organic chemistry in the 1800s changed the available forms of these drugs. Morphine was isolated in the first decade and cocaine by 1860; in 1874 diacetylmorphine was synthesized from morphine (although it became better known as heroin when the Bayer Company introduced it in 1898).

By mid-century the hypodermic syringe was perfected, and by 1870 it had become a familiar instrument to American physicians and patients [see "The Origins of Hypodermic Medication," by Norman Howard-Jones; SCIENTIFIC AMERICAN, January 1971]. At the same time, the astounding growth of the pharmaceutical industry intensified the ramifications of these accomplishments. As the century wore on, manufacturers grew increasingly adept at exploiting a marketable innovation and moving it into mass production, as well as advertising and distributing it throughout the world.

During this time, because of a peculiarity of the U.S. Constitution, the powerful new forms of opium and cocaine were more readily available in America than in most nations. Under the Constitution, individual states assumed responsibility for health issues,

DAVID F. MUSTO is professor of psychiatry at the Child Study Center and professor of the history of medicine at Yale University. He earned his medical degree at the University of Washington and received his master's in the history of science and medicine from Yale. Musto began studying the history of drug and alcohol use in the U.S. when he worked at the National Institute of Mental Health in the 1960s. He has served as a consultant for several national organizations, including the Presidential Commission on the HIV epidemic. From 1981 until 1990, Musto was a member of the Smithsonian Institution's National Council.

**HEROIN COUGH SYRUP was one of many pharmaceuticals at the turn of the century that contained mood-altering substances. The name "heroin" was coined by Bayer in 1898, a year before the company introduced aspirin.**



# COUGH

The Sum of Clinical Experience Designates Glyco-Heroin (Smith) as a Respiratory Sedative Superior in All Respects to the Preparations of Opium, Morphine, Codeine and Other Narcotics and withal devoid of the toxic or depressing effects which characterize the latter when given in doses sufficient to reduce the reflex irritability of the bronchial, tracheal and laryngeal mucous membranes.

## THE PROBLEM

of administering Heroin in proper doses in such form as will give the therapeutic virtues of this drug full sway, and will suit the palate of the most exacting adult or the most capricious child

## HAS BEEN SOLVED BY

the pharmaceutical compound known as

# GLYCO-HEROIN (Smith)

**The results attained** with GLYCO-HEROIN (SMITH) in the alleviation and cure of cough are attested by numerous clinical studies that have appeared in the medical journals within the past few years.

**Scientifically Compounded, Scientifically Conceived, GLYCO-HEROIN (SMITH) simply stands upon its merits before the profession, ready to prove its efficacy to all who are interested in the advances in the art of medication.**

### NOTES.

GLYCO-HEROIN (SMITH) is supplied to the druggist in sixteen ounce dispensing bottles only. The quantity ordinarily prescribed by the physician is two, three or four ounces.

### DOSE.

The adult dose of GLYCO-HEROIN (SMITH) is one teaspoonful, repeated every two hours or at longer intervals, as the case may require.
Children of ten or more years, from a quarter to a half teaspoonful.
Children of three years or more, five to ten drops.

SOLE BRITISH AGENTS,
THOMAS CHRISTY & CO.,
OLD SWAN LANE, UPPER THAMES ST., LONDON, E.C.

MARTIN H. SMITH & CO., Chemists,
NEW YORK CITY.

Samples and Literature Supplied on Request.

such as regulation of medical practice and the availability of pharmacological products. In fact, America had as many laws regarding health professions as it had states. For much of the 19th century, many states chose to have no controls at all; their legislatures reacted to the claims of contradictory health care philosophies by allowing free enterprise for all practitioners. The federal government limited its concern to communicable diseases and the provision of health care to the merchant marine and to government dependents.

Nations with a less restricted central government, such as Britain and Prussia, had a single, preeminent pharmacy law that controlled availability of dangerous drugs. In those countries, physicians had their right to practice similarly granted by a central authority. Therefore, when we consider consumption of opium, opiates, coca and cocaine in 19th-century America, we are looking at an era of wide availability and unrestrained advertising. The initial enthusiasm for the purified substances was only slightly affected by any substantial doubts or fear about safety, long-term health injuries or psychological dependence.

History encouraged such attitudes. Crude opium, alone or dissolved in some liquid such as alcohol, was brought by European explorers and settlers to North America. Colonists regarded opium as a familiar resource for pain relief. Benjamin Franklin regularly took laudanum—opium in alcohol extract—to alleviate the pain of kidney stones during the last few years of his life. The poet Samuel Taylor Coleridge, while a student at Cambridge in 1791, began using laudanum for pain and developed a lifelong addiction to the drug. Opium use in those early decades constituted an "experiment in nature" that has been largely forgotten, even repressed, as a result of the extremely negative reaction that followed.

Americans had recognized, however, the potential danger of continually using opium long before the availability of morphine and the hypodermic's popularity. The American Dispensatory of 1818 noted that the habitual use of opium could lead to "tremors, paralysis, stupidity and general emaciation." Balancing this danger, the text proclaimed the extraordinary value of opium in a multitude of ailments ranging from cholera to asthma. (Considering the treatments then in vogue—blistering, vomiting and bleeding—we can understand why opium was as cherished by patients as by their physicians.)

Opium's rise and fall can be tracked through U.S. import-consumption statistics compiled while importation of the drug and its derivative, morphine, was unrestricted and carried moderate tariffs. The per capita consumption of crude opium rose gradually during the 1800s, reaching a peak in the last decade of the century. It then declined, but after 1915 the data no longer reflect trends in drug use, because that year new federal laws severely restricted legal imports. In contrast, per capita consumption of smoking opium rose until a 1909 act outlawed its importation.

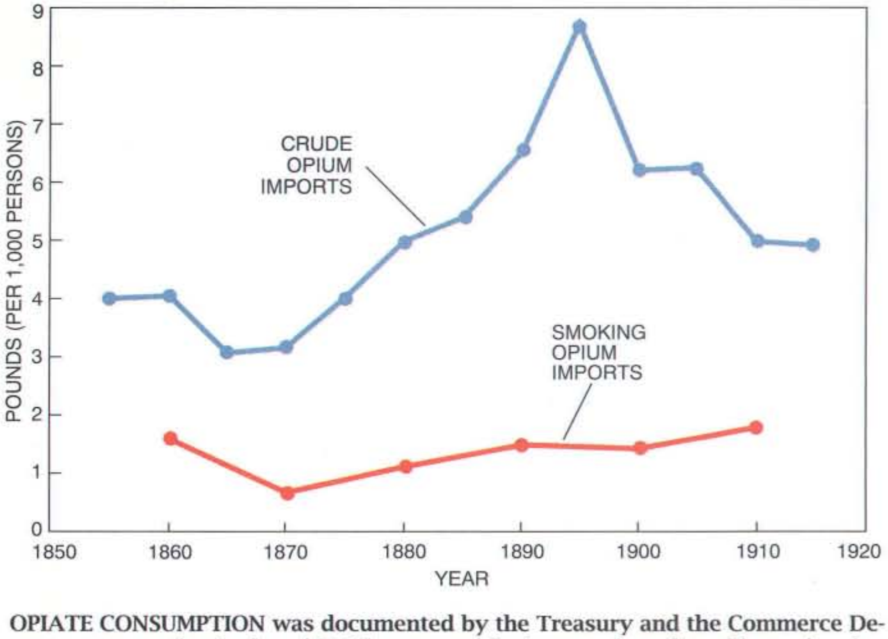
Americans had quickly associated smoking opium with Chinese immigrants who arrived after the Civil War to work on railroad construction. This association was one of the earliest examples of a powerful theme in the American perception of drugs: linkage between a drug and a feared or rejected group within society. Cocaine would be similarly linked with blacks and marijuana with Mexicans in the first third of the 20th century. The association of a drug with a racial group or a political cause, however, is not unique to America. In the 19th century, for instance, the Chinese came to regard opium as a tool and symbol of Western domination. That perception helped to fuel a vigorous antiopium campaign in China early in the 20th century.

During the 1800s, increasing numbers of people fell under the influence of opiates—substances that demanded regular consumption or the penalty of withdrawal, a painful but rarely life-threatening experience. Whatever the cause—overprescribing by physicians, over-the-counter medicines, self-indulgence or "weak will"—opium addiction brought shame. As consumption increased, so did the frequency of addiction.

At first, neither physicians nor their patients thought that the introduction of the hypodermic syringe or pure morphine contributed to the danger of addiction. On the contrary, because pain could be controlled with less morphine when injected, the presumption was made that the procedure was less likely to foster addiction.

Late in the century some states and localities enacted laws limiting morphine to a physician's prescription, and some laws even forbade refilling these prescriptions. But the absence of any federal control over interstate commerce in habit-forming drugs, of uniformity among the state laws and of effective enforcement meant that the rising tide of legislation directed at opiates—and later cocaine—was more a reflection of changing public attitude toward these drugs than an effective reduction of supplies to users. Indeed, the decline noted after the mid-1890s was probably related mostly to the public's growing fear of addiction and of the casual social use of habit-forming substances rather than to any successful campaign to reduce supplies.

At the same time, health professionals were developing more specific treatments for painful diseases, finding less



**OPIATE CONSUMPTION was documented by the Treasury and the Commerce Departments, starting in the mid-19th century. The importation of smoking opium became illegal in 1909, and crude opium and its derivatives were severely restricted in 1915. After 1915, the data reflected medicinal use.**

dangerous analgesics (such as aspirin) and beginning to appreciate the addictive power of the hypodermic syringe. By now the public had learned to fear the careless, and possibly addicted, physician. In *A Long Day's Journey into Night,* Eugene O'Neill dramatized the painful and shameful impact of his mother's physician-induced addiction.

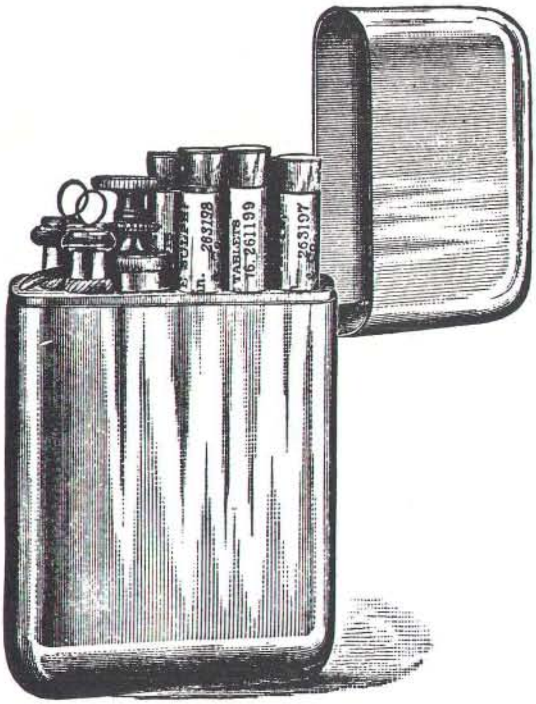
In a spirit not unlike that of our times, Americans in the last decade of the 19th century grew increasingly concerned about the environment, adulterated foods, destruction of the forests and the widespread use of mood-altering drugs. The concern embraced alcohol as well. The Anti-Saloon League, founded in 1893, led a temperance movement toward prohibition, which later was achieved in 1919 and became law in January 1920.

After overcoming years of resistance by over-the-counter, or patent, medicine manufacturers, the federal government enacted the Pure Food and Drug Act in 1906. This act did not prevent sales of addictive drugs like opiates and cocaine, but it did require accurate labeling of contents for all patent remedies sold in interstate commerce. Still, no national restriction existed on the availability of opiates or cocaine. The solution to this problem would emerge from growing concern, legal ingenuity and the unexpected involvement of the federal government with the international trade in narcotics.

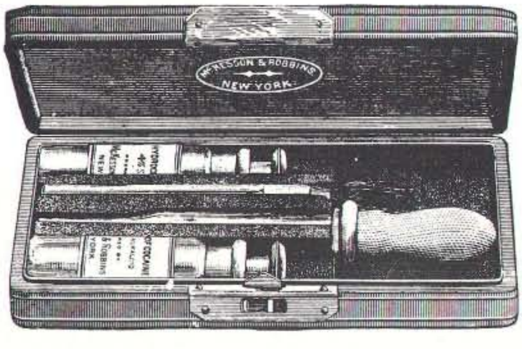
Responsibility for the Philippines in 1898 added an international dimension to the growing domestic alarm about drug abuse. It also revealed that Congress, if given the opportunity, would prohibit nonmedicinal uses of opium among its new dependents. Civil Governor William Howard Taft proposed reinstituting an opium monopoly—through which the previous Spanish colonial government had obtained revenue from sales to opium merchants—and using those profits to help pay for a massive public education campaign. President Theodore Roosevelt vetoed this plan, and in 1905 Congress mandated an absolute prohibition of opium for any purpose other than medicinal use.

To deal efficiently with the antidrug policy established for the Philippines, a committee from the Islands visited various territories in the area to see how others dealt with the opium problem. The benefit of controlling narcotics internationally became apparent.

In early 1906 China had instituted a campaign against opium, especially smoking opium, in an attempt to modernize and to make the Empire better able to cope with continued Western encroachments on its sovereignty. At about the same time, Chinese anger at maltreatment of their nationals in the U.S. seethed into a voluntary boycott of American goods. Partly to appease the Chinese by aiding their antiopium efforts and partly to deal with uncontrollable smuggling within the Philippine Archipelago, the U.S. convened a meeting of regional powers. In this way, the U.S. launched a campaign for worldwide narcotics traffic control that would extend through the years in an unbroken diplomatic sequence from the League of Nations to the present efforts of the United Nations.



**1894 EMERGENCY KIT by the Parke-Davis Company carried cocaine, morphine, atropine and strychnine as well as a hypodermic syringe.**



**POCKET COCAINE CASE manufactured by pharmacists McKesson & Robbins was one of many drug kits on the market in the late 1800s.**

The International Opium Commission, a gathering of 13 nations, met in Shanghai in February 1909. The Protestant Episcopal bishop of the Philippines, Charles Henry Brent, who had been instrumental in organizing the meeting, was chosen to preside. Resolutions noting problems with opium and opiates were adopted, but they did not constitute a treaty, and no decisions bound the nations attending the commission. In diplomatic parlance, what was needed now was a conference not a commission. The U.S. began to pursue this goal with determination.

The antinarcotics campaign in America had several motivations. Appeasement of China was certainly one factor for officials of the State Department. The department's opium commissioner, Hamilton Wright, thought the whole matter could be "used as oil to smooth the troubled water of our aggressive commercial policy there." Another reason was the belief, strongly held by the federal government today, that controlling crops and traffic in producing countries could most efficiently stop U.S. nonmedical consumption of drugs.

To restrict opium and coca production required worldwide agreement and, thus, an international conference. After intense diplomatic activity, one was convened in the Hague in December 1911. Brent again presided, and on January 23, 1912, the 12 nations represented signed a convention. Provision was made for the other countries to comply before the treaty was brought into force. After all, no producing or manufacturing nation wanted to leave the market open to nonratifying nations.

The convention required each country to enact domestic legislation controlling narcotics trade. The goal was a world in which narcotics were restricted to medicinal use. Both the producing and consuming nations would have control over their boundaries.

After his return from Shanghai, Wright labored to craft a comprehensive federal antinarcotics law. In his path loomed the problem of states' rights. The health professions were considered a major cause of patient addiction. Yet how could federal law interfere with the prescribing practices of physicians or require that pharmacists keep records? Wright settled on the federal government's power to tax; the result, after prolonged bargaining with pharmaceutical, import, export and medical interests, was the Harrison Act of December 1914.

Representative Francis Burton Harrison's association with the act was an accidental one, the consequence of his introduction of the administration's bill. If the chief proponent and negotiator were to be given eponymic credit, it should have been called the Wright Act. It could even have been called a second Mann Act, after Representative James Mann, who saw the bill through to passage in the House of Representatives, for by that time Harrison had become governor-general of the Philippines.

The act required a strict accounting of opium and coca and their derivatives from entry into the U.S. to dispensing to a patient. To accomplish this control, a small tax had to be paid at each transfer, and permits had to be obtained by applying to the Treasury Department. Only the patient paid no tax, needed no permit and, in fact, was not allowed to obtain one.

Initially Wright and the Department of Justice argued that the Harrison Act forbade indefinite maintenance of addiction unless there was a specific medical reason such as cancer or tuberculosis. This interpretation was rejected in 1916 by the Supreme Court—even though the Justice Department argued that the Harrison Act was the domestic implementation of the Hague Opium Convention and therefore took precedence over states' rights. Maintenance was to be allowed.

That decision was short-lived. In 1919 the Supreme Court, led by Oliver Wendell Holmes and Louis Brandeis, changed its mind by a 5-4 vote. The court declared that indefinite maintenance for "mere addiction" was outside legitimate medical practice and that, consequently, prohibiting it did not constitute interference with a state's right to regulate physicians. Second, because the person receiving the drugs for maintenance was not a bona fide patient but just a recipient of drugs, the transfer of narcotics defrauded the government of taxes required under the Harrison Act.

During the 1920s and 1930s, the opiate problem, chiefly morphine and heroin, declined in the U.S., until much of the problem was confined to the periphery of society and the outcasts of urban areas. There were exceptions: some health professionals and a few others of middle class or higher status continued to take opiates.

**CHARLES HENRY BRENT, the Protestant Episcopal bishop of the Philippines, chaired the first two international meetings on opium control.**

America's international efforts continued. After World War I, the British and U.S. governments proposed adding the Hague Convention to the Versailles Treaty. As a result, ratifying the peace treaty meant ratifying the Hague Convention and enacting a domestic law controlling narcotics. This incorporation led to the British Dangerous Drugs Act of 1920, an act often misattributed to a raging heroin epidemic in Britain. In the 1940s some Americans argued that the British system provided heroin to addicts and, by not relying on law enforcement, had almost eradicated the opiate problem. In fact, Britain had no problem to begin with. This argument serves as an interesting example of how the desperate need to solve the drug problem in the U.S. tends to create misperceptions of a foreign drug situation.

The story of cocaine use in America is somewhat shorter than that of opium, but it follows a similar plot. In 1884 purified cocaine became commercially available in the U.S. At first the wholesale cost was very high—$5 to $10 a gram—but it soon fell to 25 cents a gram and remained there until the price inflation of World War I. Problems with cocaine were evident almost from the beginning, but popular opinion and the voices of leading medical experts depicted cocaine as a remarkable, harmless stimulant.

William A. Hammond, one of America's most prominent neurologists, extolled cocaine in print and lectures. By 1887 Hammond was assuring audiences that cocaine was no more habit-forming than coffee or tea. He also told them of the "cocaine wine" he had perfected with the help of a New York druggist: two grains of cocaine to a pint of wine. Hammond claimed that this tonic was far more effective than the popular French coca wine, probably a reference to Vin Mariani, which he complained had only half a grain of cocaine to the pint.

Coca-Cola was also introduced in 1886 as a drink offering the advantages of coca but lacking the danger of alcohol. It amounted to a temperance coca beverage. The cocaine was removed in 1900, a year before the city of Atlanta, Ga., passed an ordinance (and a state statute the following year) prohibiting provision of any cocaine to a consumer without a prescription.

Cocaine is one of the most powerful of the central nervous system euphoriants. This fact underlay cocaine's quickly growing consumption and the ineffectiveness of the early warnings. How could anything that made users so confident and happy be bad? Within a year of cocaine's introduction, the Parke-Davis Company provided coca and cocaine in 15 forms, including coca cigarettes, cocaine for injection and cocaine for sniffing. Parke-Davis and at least one other company also offered consumers a handy cocaine kit. (The Parke-Davis kit contained a hypodermic syringe.) The firm proudly supplied a drug that, it announced, "can supply the place of food, make the coward brave, the silent eloquent and... render the sufferer insensitive to pain."

Cocaine spread rapidly throughout the nation. In September 1886 a physician in Puyallup, Washington Territory, reported an adverse reaction to cocaine during an operation. Eventually reports of overdoses and idiosyncratic reactions shifted to accounts of the social and behavioral effects of long-term cocaine use. The ease with which experimenters became regular users and the increasing instances of cocaine being linked with violence and paranoia gradually took hold in popular and medical thought.

In 1907 an attempt was made in New York State to shift the responsibility for cocaine's availability from the open market to medical control. Assemblyman Alfred E. Smith, later the governor of New York and in 1928 the Democratic party's presidential candidate, sponsored such a bill. The cost of cocaine on New York City streets, as revealed by newspaper and police accounts after the law's enactment, was typically 25 cents a packet, or "deck."

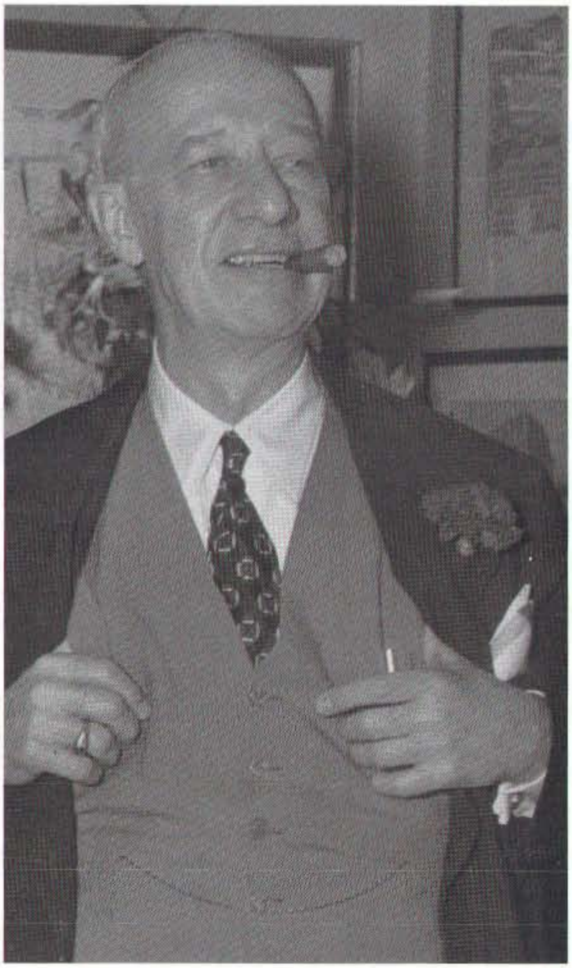
Although 25 cents may seem cheap, it was actually slightly higher than the average industrial wage at that time, which was about 20 cents an hour. Packets, commonly glycine envelopes, usually contained one to two grains (65 to 130 milligrams), or about a tenth of a gram. The going rate was roughly 10 times that of the wholesale price, a ratio not unlike recent cocaine street prices, although in the past few years the street price has actually been lower in real value than what it was in 1910.

Several similar reports from the years before the Harrison Act of 1914 suggest that both the profit margin and the street price of cocaine were unaffected by the legal availability of cocaine from a physician. Perhaps the formality of medical consultation and the growing antagonism among physicians and the public toward cocaine helped to sustain the illicit market.

In 1910 William Howard Taft, then president of the U.S., sent to Congress a report that cocaine posed the most serious drug problem America had ever faced. Four years later President Woodrow Wilson signed into law the Harrison Act, which, in addition to its opiate provisions, permitted the sale of cocaine only through prescriptions. It also forbade any trace of cocaine in patent remedies, the most severe restriction on any habit-forming drug to that date. (Opiates, including heroin, could still be present in small amounts in nonprescription remedies, such as cough medicines.)

Although the press continued to reveal Hollywood scandals and underworld cocaine practices during the 1920s, cocaine use gradually declined as a societal problem. The laws probably hastened the trend, and certainly the tremendous public fear reduced demand. By 1930 the New York City Mayor's Committee on Drug Addiction was reporting that "during the last 20 years cocaine as an addiction has ceased to be a problem."

Unlike opiates and cocaine, marijuana was introduced during a period of drug intolerance. Consequently, it was not until the 1960s, 40 years after marijuana cigarettes had arrived in America, that it was widely used. The practice of smoking cannabis leaves came to the U.S. with Mexican immigrants, who had come North during the 1920s to work in agriculture,



**STATESMEN AND DRUGS: William Howard Taft (*left*) dealt with the drug problem twice—first as civil governor of the Philippines and then as president of the U.S. Alfred E. Smith (*right*), New York State assemblyman and governor and the 1928 Democratic presidential candidate, sponsored a state bill to control the availability of cocaine.**

and it soon extended to white and black jazz musicians.

As the Great Depression of the 1930s settled over America, the immigrants became an unwelcome minority linked with violence and with growing and smoking marijuana. Western states pressured the federal government to control marijuana use. The first official response was to urge adoption of a uniform state antinarcotics law. Then a new approach became feasible in 1937, when the Supreme Court upheld the National Firearms Act. This act prohibited the transfer of machine guns between private citizens without purchase of a transfer tax stamp—and the government would not issue the necessary stamp. Prohibition was implemented through the taxing power of the federal government.

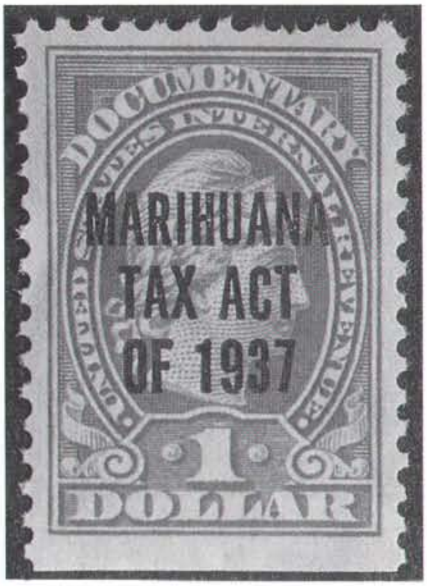
Within a month of the Supreme Court's decision, the Treasury Department testified before Congress for a bill to establish a marijuana transfer tax. The bill became law, and until the Comprehensive Drug Abuse Act of 1970, marijuana was legally controlled through a transfer tax for which no stamps or licenses were available to private citizens. Certainly some people were smoking marijuana in the 1930s, but not until the 1960s was its use widespread.

Around the time of the Marihuana Tax Act of 1937, the federal government released dramatic and exaggerated portrayals of marijuana's effects. Scientific publications during the 1930s also fearfully described marijuana's dangers. Even Walter Bromberg, who thought that marijuana made only a small contribution to major crimes, nevertheless reported the drug was "a primary stimulus to the impulsive life with direct expression in the motor field."

Marijuana's image shifted during the 1960s, when it was said that its use at the gigantic Woodstock gathering kept peace—as opposed to what might have happened if alcohol had been the drug of choice. In the shift to drug toleration in the late 1960s and early 1970s, investigators found it difficult to associate health problems with marijuana use. The 1930s and 1940s had marked the nadir of drug toleration in the U.S., and possibly the mood of both times affected professional perception of this controversial plant.

After the Harrison Act, the severity of federal laws concerning the sale and possession of opiates and cocaine gradually rose. As drug use declined, penalties increased until 1956, when the death penalty was introduced as an option by the federal government for anyone older than 18 providing heroin to anyone younger than 18 (apparently no one was ever executed under this statute). At the same time, mandatory minimum prison sentences were extended to 10 years.

After the youthful counterculture discovered marijuana in the 1960s, demand for the substance grew until about 1978, when the favorable attitude toward it reached a peak. In 1972 the Presidential Commission on Marihuana and Drug Abuse recommended "decriminalization" of marijuana, that is, legal possession of a small amount for personal use. In 1977 the Carter administration formally advocated legalizing marijuana in amounts up to an ounce.



**MARIHUANA TAX STAMP of 1937 established governmental control over the transfer and sale of the plant. The stamp was never available for private use.**

The Gallup Poll on relaxation of laws against marijuana is instructive. In 1980, 53 percent of Americans favored legalization of small amounts of marijuana; by 1986 only 27 percent supported that view. At the same time, those favoring penalties for marijuana use rose from 43 to 67 percent. This reversal parallels the changes in attitude among high school students revealed by the Institute of Social Research at the University of Michigan.

The decline in favorable attitudes toward marijuana that began in the late 1970s continues. In the past few years we have seen penalties rise again against users and dealers. The recriminalization of marijuana possession by popular vote in Alaska in 1990 is one example of such a striking reversal.

In addition to stricter penalties, two other strategies, silence and exaggeration, were implemented in the 1930s to keep drug use low and prevent a recurrence of the decades-long, frustrating and fearful antidrug battle of the late 19th and early 20th centuries. Primary and secondary schools instituted educational programs against drugs. Then policies shifted amid fears that talking about cocaine or heroin to young people, who now had less exposure to drugs, would arouse their curiosity. This concern led to a decline in drug-related information given during school instruction as well as to the censorship of motion pictures.

The Motion Picture Association of America, under strong public and religious pressure, decided in 1934 to refuse a seal of approval for any film that showed narcotics. This prohibition was enforced with one exception—*To the Ends of the Earth,* a 1948 film that lauded the Federal Bureau of Narcotics—until *Man with a Golden Arm* was successfully exhibited in 1956 without a seal.

Associated with a decline in drug information was a second, apparently paradoxical strategy: exaggerating the effects of drugs. The middle ground was abandoned. In 1924 Richmond P. Hobson, a nationally prominent campaigner against drugs, declared that one ounce of heroin could addict 2,000 persons. In 1936 an article in the *American Journal of Nursing* warned that a marijuana user "will suddenly turn with murderous violence upon whomever is nearest to him. He will run amuck with knife, axe, gun, or anything else that is close at hand, and will kill or maim without any reason."

A goal of this well-meaning exaggeration was to describe drugs so repulsively that anyone reading or hearing of them would not be tempted to experiment with the substances. One contributing factor to such a publicity campaign, especially regarding marijuana, was that the Depression permitted little money for any other course of action.

Severe penalties, silence and, if silence was not possible, exaggeration became the basic strategies against drugs after the decline of their first wave of use. But the effect of these tactics was to create ignorance and false images that would present no real obstacle to a renewed enthusiasm for drugs in the 1960s. At the time, enforcing draconian and mandatory penalties would have filled to overflowing all jails and prisons with the users of marijuana alone.

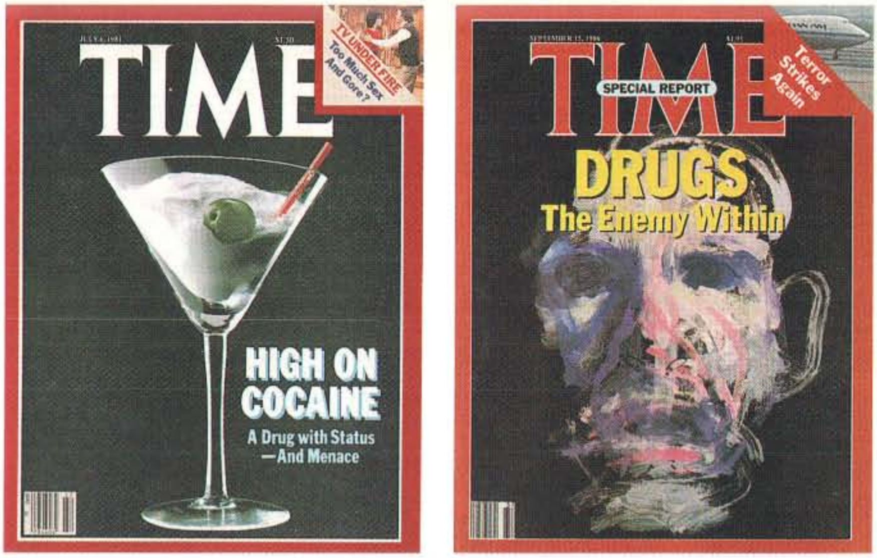
Exaggeration fell in the face of the realities of drug use and led to a loss of credibility regarding any government pronouncement on drugs. The lack of information erased any awareness of the first epidemic, including the gradually obtained and hard-won public insight into the hazards of cocaine and opiates. Public memory, which would have provided some context for the antidrug laws, was a casualty of the antidrug strategies.

The earlier and present waves of drug use have much in common, but there is at least one major difference. During the first wave of drug use, antidrug laws were not enacted until the public demanded them. In contrast, today's most severe antidrug laws were on the books from the outset; this gap between law and public opinion made the controls appear ridiculous and bizarre. Our current frustration over the laws' ineffectiveness has been greater and more lengthy than before because we have lived through many years in which antidrug laws lacked substantial public support. Those laws appeared powerless to curb the rise in drug use during the 1960s and 1970s.

The first wave of drug use involved primarily opiates and cocaine. The nation's full experience with marijuana is now under way (marijuana's tax regulation in 1937 was not the result of any lengthy or broad experience with the plant). The popularity and growth in demand for opiates and cocaine in mainstream society derived from a simple factor: the effect on most people's physiology and emotions was enjoyable. Moreover, Americans have recurrently hoped that the technology of drugs would maximize their personal potential. That opiates could relax and cocaine energize seemed wonderful opportunities for fine-tuning such efforts.

Two other factors allowed a long and substantial rise in consumption during the 1800s. First, casualties accumulate gradually; not everyone taking cocaine or opiates becomes hooked on the drug. In the case of opiates, some users have become addicted for a lifetime and have still been productive.

Yet casualties have mounted as those who could not handle occasional use have succumbed to domination by drugs and by drug-seeking behavior. These addicts become not only miserable themselves but also frightening to their families and friends. Such cases are legion today in our larger cities, but the percentage of those who try a substance and acquire a dependence or get into serious legal trouble is not 100 percent. For cocaine, the estimate varies from 3 to 20 percent, or even higher, and so it is a matter of time before cocaine is recognized as a likely danger.



***TIME* MAGAZINE COVERS from 1981 and from 1986 reflect a clear change in American views toward mood-altering drugs, specifically cocaine.**

Early in the cycle, when social tolerance prevails, the explanation for casualties is that those who succumb to addiction are seen as having a physiological idiosyncrasy or "foolish trait." Personal disaster is thus viewed as an exception to the rule. Another factor minimizing the sense of risk is our belief in our own invulnerability—that general warnings do not include us. Such faith reigns in the years of greatest exposure to drug use, ages 15 to 25. Resistance to a drug that makes a user feel confident and exuberant takes many years to permeate a society as large and complex as the U.S.

The interesting question is not why people take drugs, but rather why they stop taking them. We perceive risk differently as we begin to reject drugs. One can perceive a hypothetical 3 percent risk from taking cocaine as an assurance of 97 percent safety, or one can react as if told that 3 percent of New York/Washington shuttle flights crash. Our exposure to drug problems at work, in our neighborhood and within our families shifts our perception, gradually shaking our sense of invulnerability.

Cocaine has caused the most dramatic change in estimating risk. From a grand image as the ideal tonic, cocaine's reputation degenerated into that of the most dangerous of drugs, linked in our minds with stereotypes of mad, violent behavior. Opiates have never fallen so far in esteem, nor were they repressed to the extent cocaine had been between 1930 and 1970.

Today we are experiencing the reverse of recent decades, when the technology of drug use promised an extension of our natural potential. Increasingly we see drug consumption as reducing what we could achieve on our own with healthy food and exercise. Our change of attitude about drugs is connected to our concern over air pollution, food adulteration and fears for the stability of the environment.

Ours is an era not unlike that early in this century, when Americans made similar efforts at self-improvement accompanied by an assault on habit-forming drugs. Americans seem to be the least likely of any people to accept the inevitability of historical cycles. Yet if we do not appreciate our history, we may again become captive to the powerful emotions that led to draconian penalties, exaggeration or silence.

FURTHER READING

AMERICAN DIPLOMACY AND THE NARCOTICS TRAFFIC, 1900–1939. Arnold H. Taylor. Duke University Press, 1969.

DRUGS IN AMERICA: A SOCIAL HISTORY, 1800–1980. H. Wayne Morgan. Syracuse University Press, 1981.

DARK PARADISE: OPIATE ADDICTION IN AMERICA BEFORE 1940. David T. Courtwright. Harvard University Press, 1982.

THE AMERICAN DISEASE: ORIGINS OF NARCOTIC CONTROL. Expanded Edition. David F. Musto. Oxford University Press, 1987.

AMERICA'S FIRST COCAINE EPIDEMIC. David F. Musto in *Wilson Quarterly*, pages 59–65; Summer 1989.

ILLICIT PRICE OF COCAINE IN TWO ERAS: 1908–14 AND 1982–89. David F. Musto in *Connecticut Medicine*, Vol. 54, No. 6, pages 321–326; June 1990.

# The Early Life of Stars

*Violent collisions of gas dominate the childhood of stars. Eventually nuclear fusion begins, enabling mature stars to burn steadily for billions of years*

by Steven W. Stahler

Looking up at a clear night sky, far from city lights, one sees that the universe is filled with stars. Somehow nature has managed to create untold numbers of these objects—an estimated 100 billion in the Milky Way alone. Furthermore, stars continue to be born, 10 to 20 billion years after the universe began. How are stars created? What changes does a young star undergo before it settles into the relatively stable state now exhibited by the sun?

From a physicist's perspective, a star is a ball of hot gas held together by its own gravity. The heat and pressure generated by internal nuclear reactions—chiefly the fusion of hydrogen into helium—prevent the star from collapsing under the gravitational force. This relatively simple system has a well-defined life span. It begins with the star's condensation from a diffuse cloud of interstellar gas and ends when the star, having exhausted its nuclear fuel, fades from sight as a white dwarf, neutron star or black hole.

From this description, it might seem that detailing the formation and early evolution of stars should present no essential difficulties. But the subtle ways in which gravity and thermal pressure interact cause young stars to behave in a manner that sometimes defies intuition. Consider, for example, the evolution of luminosity, the amount of energy emitted by the stellar surface per unit time. The internal temperature of a young star is too low to fuse hydrogen, so its luminosity would also be expected to be relatively low. It might increase when the fusion of hydrogen begins and then gradually fade.

STEVEN W. STAHLER received his Ph.D. from the University of California, Berkeley, in 1980 and has held positions at Cornell University and the Harvard-Smithsonian Center for Astrophysics. Since 1985, he has been professor of physics at the Massachusetts Institute of Technology. Stahler writes that "this is an exhilarating time" for the study of young stars, likening the current state of research to being in a darkroom as the image gradually appears—an apt metaphor, for he also pursues an avid interest in filmmaking. With Mary Barsony, he is writing *The Formation of Stars,* to be published by Addison-Wesley in 1992.

In fact, a very young star is extremely bright. Its luminosity decreases as age advances, reaching a temporary minimum at the time of hydrogen ignition. The early life of stars involves a rich variety of physical processes, some of which remain poorly understood. Only within the past two decades have astronomers begun to piece together, through theory and observation, a coherent and detailed picture.

Stars condense under their own gravity from large, optically invisible clouds found throughout the disks of spiral galaxies. Such aggregates are called giant molecular cloud complexes. The term "molecular" refers to the fact that the gas consists mostly of hydrogen in its molecular form [see "Giant Molecular-Cloud Complexes in the Galaxy," by Leo Blitz; SCIENTIFIC AMERICAN, April 1982]. The complexes are the most massive structures in the galaxy, sometimes measuring more than 300 light-years across.

Closer inspection reveals that stars develop from isolated condensations within the giant molecular cloud complexes. Such condensations are called dense cores. Philip C. Myers of the Harvard-Smithsonian Center for Astrophysics, who coined the term in 1983, was the first to observe their properties systematically and to emphasize their role in star formation.

Astronomers investigate the properties of dense cores by using large radio telescopes, the only instruments capable of detecting the weak, millimeter-wavelength radiation that the clouds emit. The radiation comes not from the molecular hydrogen but from the trace amounts of other substances in the cores, such as carbon monoxide and carbon monosulfide. The emissions from these trace gases reveal that a typical dense core has a diameter of a few light-months, a density of 30,000 hydrogen molecules per cubic centimeter, and a temperature of 10 kelvins [*see illustration on page 33*].

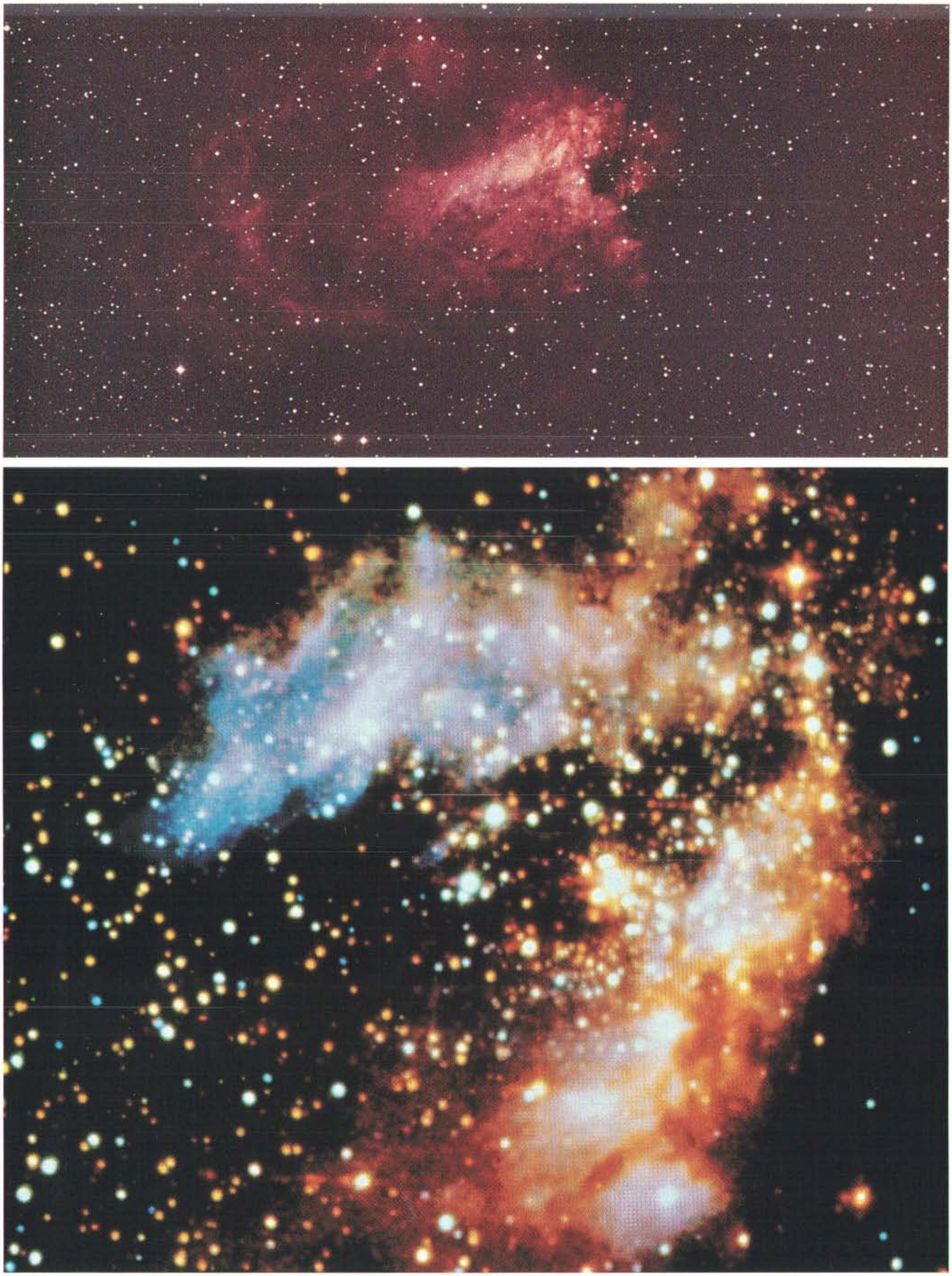
From these numbers, investigators inferred that the gas pressure in a dense core is just about the right magnitude to withstand the compressive force of the core's own gravity. To form a star, therefore, the core must collapse from a marginally unstable state—that is, one in which gravity is only slightly stronger than pressure.

How the core itself condensed from its parent molecular cloud complex to arrive at this marginally unstable state is still not well understood. Nevertheless, astrophysicists had the tools to model stellar formation even before the discovery of dense cores. In the 1960s theorists had used computer simulations to determine how clouds in unstable states collapse.

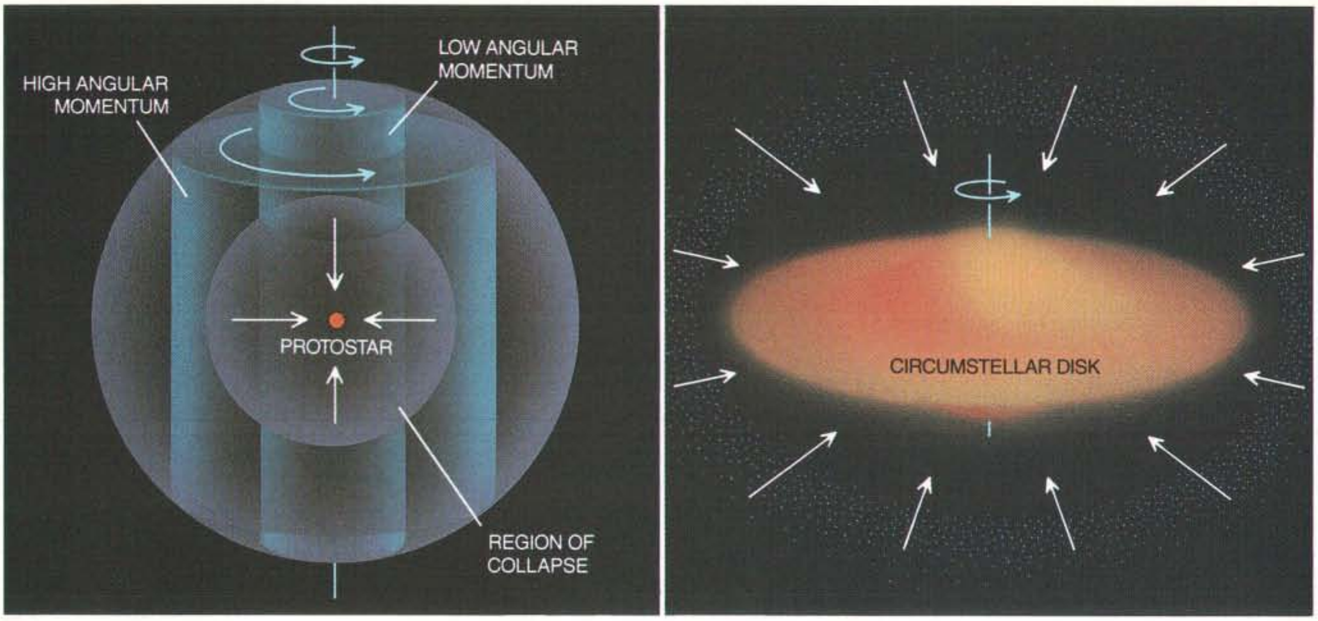
Although the simulations assumed widely varying initial conditions, each one showed that clouds that are not violently unstable collapse in an inside-out manner. That is, material at the center first enters into a true free-fall collapse while the outlying gas remains static. Gradually the region of collapse spreads outward through the rest of the cloud.

Deep within the collapsing region, a star begins to form from the collisions of gas. The star itself is only about one light-second in diameter—one ten-millionth that of the dense core. For a star of such a relatively small size, the overall pattern of collapse is insignificant. What counts is the mass accretion rate.

This rate is the amount of matter per unit time crossing an imaginary spherical shell near the center of the cloud. In his elegant and influential 1977 paper, Frank H. Shu of the University of California at Berkeley demonstrated a remarkable result: the mass-accretion rate depends solely on the initial cloud



**OMEGA NEBULA** in Sagittarius (*optical image at top*) is a region where stars form. This stellar nursery shows little of its interior in visible light, but an infrared image of the right-hand part of the nebula reveals many young stars (*bottom*).



**DENSE CORES collapse from the inside out. In a highly idealized model (*left*), gas within the inner sphere falls onto the protostar surface. If the dense core rotates uniformly, angular momentum remains constant on cylinders centered on the rotation axis. The region of collapse expands to engulf material with higher angular momentum; this material misses the protostar and orbits it as a disk (*right*). The protostar itself becomes rotationally distorted.**

temperature. The higher the temperature, the greater the rate of accretion. Shu's result indicates that an amount of mass equal to the sun's accumulates at the center of a collapsing dense core in 100,000 to one million years.

The object forming at the center of the collapsing cloud is known as a protostar. The modern theory of protostars began to take shape in 1969, when Richard Larson of Yale University was able to observe stellar buildup in a computer simulation of cloud collapse. Researchers developing Larson's pioneering work have found an advantage in conceptually separating the protostar from the cloud as they model the collapse. In other words, investigators examine the protostar in isolation as an ordinary star with an extraordinary outer boundary condition—the incoming accretion flow.

Astronomers running these simulations can vary the properties of the accretion flow to gauge its effect on the evolution of the protostar. In 1980 Shu, Ronald Taam of Northwestern University and I first used this approach to establish the properties of protostars of about one solar mass. More recently, in collaboration with Francesco Palla of Arcetri Observatory in Florence, I have turned to the method once again to study more massive protostars.

Through such simulations, astronomers have now developed a model that describes the protostar phase. Workers have found that the incoming gas hits the protostar at a very high speed, so high that it cannot decelerate gradually before reaching the stellar surface. Instead it encounters a strong shock front (a sharp transition to very high pressure) that rapidly halts the gas. Within the shock, the gas is heated to nearly one million kelvins. The gas quickly cools by radiation to about 10,000 kelvins and settles down, layer by layer, to form the protostar.

The shock front explains why young stars are so luminous. If a protostar attains one solar mass, the luminosity that the gas generates when it encounters the shock front exceeds solar brightness by a factor of six to 60. Thus, the acute brightness of these young stars stems not from nuclear fusion, as in ordinary stars, but from the kinetic energy of matter as it is pulled in by gravity.

The luminosity from protostars is observable, but not with optical telescopes. All the gas in interstellar space, including that which forms stars, contains "dust," a mixture of solid particles of submicron size. As photons stream outward from the shock front, they eventually encounter massive amounts of these dust grains falling in along with the gas of the original dense core.

The dust cannot reach the protostar surface, because the intense heat from the shock front vaporizes it. Astronomers refer to the volume of space within which the dust is vaporized as the opacity gap. Farther upstream, beyond the opacity gap, temperatures are low enough to enable the grains to exist. The cold grains absorb the shock-generated photons and reemit them at longer wavelengths. These long-wavelength photons are in turn absorbed by dust grains lying farther away.

The photons thus tortuously make their way through the cloud material, until their average wavelength lies deep in the infrared region of the electromagnetic spectrum. At a radius that Shu, Taam and I call the dust photosphere, located a few light-hours from the protostar, the photons have too long a wavelength to be absorbed by the dust and can finally fly unimpeded to earth-based infrared telescopes.

Despite the capabilities of modern detectors, astronomers cannot definitively say that telescopes have actually recorded the infrared signals of protostars. After its launch in 1983, the *Infrared Astronomical Satellite* generated hundreds of thousands of images of pointlike infrared sources of radiation [see "The Infrared Sky," by Harm J. Habing and Gerry Neugebauer; SCIENTIFIC AMERICAN, November 1984]. Many sources appear to be located deep within the radio-imaged dense cores; some must undoubtedly be protostars. The uncertainty arises because detectors cannot distinguish the protostars from somewhat older stars, also buried in dust and gas.

For a positive identification, infrared or radio telescopes must be able to detect the Doppler shift of spectral lines very close to an infrared point source.

The Doppler shift would represent the actual motion of the gas as it falls onto the stellar surface.

Once the protostar accretes sufficient material to reach a few tenths the mass of the sun, the temperature at the center becomes sufficient to induce nuclear fusion. Fusion in protostars, however, is quite different from that of main-sequence stars—that is, middle-aged stars like the sun existing in a long-lived state of equilibrium. The primary reaction that powers a mature star involves the fusion of hydrogen nuclei.

Hydrogen is the most common chemical constituent of the universe. The big bang created this element primarily in its normal isotopic form, as an atom whose nucleus consists of a single proton. But about two out of every 100,000 hydrogen nuclei consist of deuterium, a nucleus of one proton and one neutron. Deuterium persists today in the interstellar gas that becomes incorporated in new stars.

Remarkably, this tiny impurity plays a dominant role in the life of protostars. The interiors of protostars are not yet hot enough to fuse ordinary hydrogen, a reaction that occurs at about 10 million kelvins. But protostars can easily attain, as a result of the compressive force of gravity, the temperature of one million kelvins required to initiate the fusion of deuterium, which also liberates large amounts of energy. The protostellar material is too opaque to transmit this energy by radiation. Instead the star becomes convectively unstable: bubbles of gas heated by the nuclear fire rise up toward the surface.

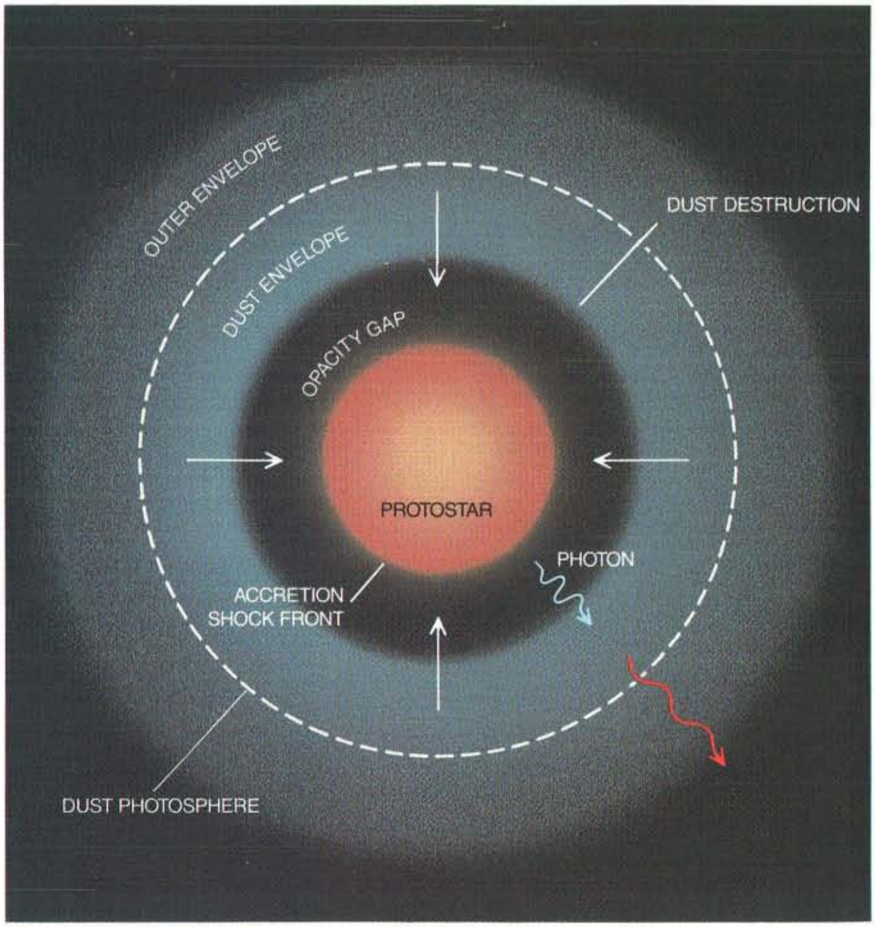
This upward motion is balanced by the sinking of cooler gas toward the center. The same sort of convective circulation, on a much smaller scale, occurs in the air in a radiator-heated room. But in a protostar, the circulating eddies drag down fresh deuterium that has landed on the surface. These deuterium atoms are quickly transported to the center, where they fuse together and release more heat. Thus, the return stroke of the convection cycle continuously resupplies the fuel needed to maintain both burning and convection.

If the protostar gains enough matter to become about twice as massive as the sun, the convection cycle begins to operate in a slightly different manner. Palla and I recently found that a thin shell of gas in the interior region becomes transparent enough to transport heat through radiation rather than convection. Neither rising nor sinking gas can penetrate this radiative barrier. Consequently, fusion quickly consumes all the deuterium inside the barrier. Fresh deuterium falling onto the protostar piles up on its surface. The compressed surface layers become hotter until they, too, ignite the deuterium, which then burns in a shell overlying the depleted interior. Hot bubbles rise up from this burning shell, make their way to the surface and then sink back down to the shell, completing the refueling cycle.

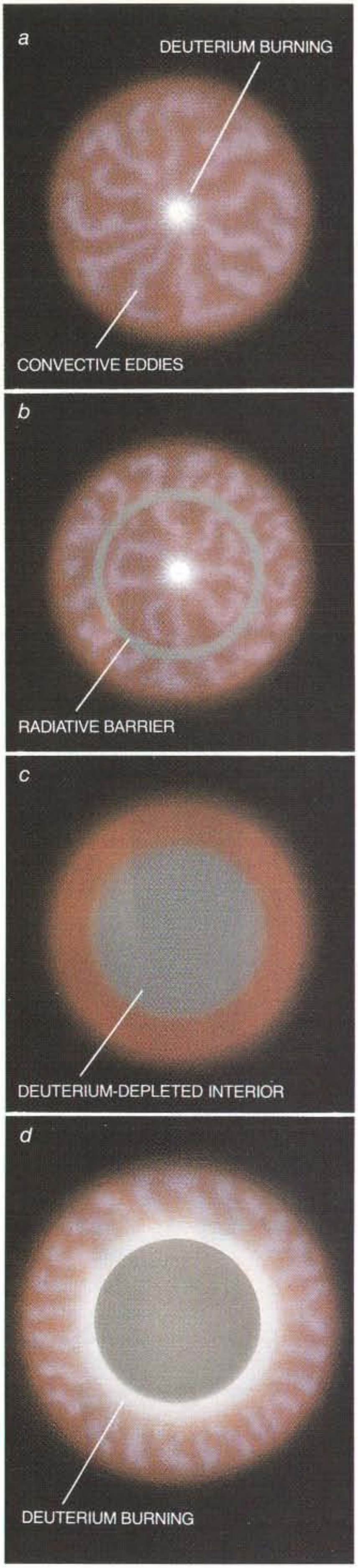
Despite the small concentration of deuterium nuclei, the heat released by their fusion has a considerable impact on the protostar. The chief effect of the burning of deuterium is to cause the protostar to swell. Because convection efficiently spreads the heat, deuterium burning enlarges each protostar to a characteristic size, determined by the object's mass. A protostar of one solar mass has a radius five times that of the present sun. A protostar of three solar masses, in which deuterium burns in a subsurface shell, swells even more dramatically: its radius is 10 times the solar value.

A typical dense core encompasses more mass than the star it ultimately produces. Therefore, some mechanism must expel the extra mass and halt the accretion. Most astronomers have now become convinced that a strong wind erupting at the surface of the protostar is responsible. The wind blows back the incoming gas and eventually disperses the entire dense core.

The wind idea did not come from any theoretical calculation. The phenomenon was suggested to surprised theorists by widespread observation of molecular gas streaming away from infrared sources of radiation [see "Energetic Outflows from Young Stars," by Charles J. Lada; SCIENTIFIC AMERICAN, July 1982]. The agent of the outflow would appear to be the protostellar wind. This wind, which has not yet been



**PROTOSTAR interacts strongly with the incoming matter that forms it, giving rise to distinct zones. Material from the outer envelope rams into the stellar surface, creating an accretion shock. The photons from the shock first fly away from the protostar. They next pass through the opacity gap, a region so hot that it vaporizes dust grains. When the photons reach the dust envelope, they are continuously absorbed and reemitted by the submicron-size particles there. Ultimately, the photons escape as infrared radiation at the dust photosphere.**



**DEUTERIUM FUSION occurs in a manner that depends on the mass of the protostar. In low-mass protostars, deuterium is fed to the burning center by turbulent convective eddies (*a*). If the protostar continues to gain mass, a radiative barrier appears, cutting off the supply of fresh deuterium to the center (*b*). The inner region quickly scours out its deuterium and reverts to a stable state, with no convection (*c*). If the protostar continues to accrete matter, deuterium will ignite in a thick shell and expand the protostar (*d*).**

directly observed, must drive off matter and energy at a vastly greater rate than do winds emanating from main-sequence stars. The cause of the protostellar wind is one of the deepest mysteries in the study of young stars.

Once the dense core disperses, the exposed object, now optically visible, is known as a pre-main-sequence star. Like a protostar, a pre-main-sequence star is highly luminous. Once again, gravity rather than nuclear fusion accounts for the luminosity. The pressure within the star prevents it from going into a true free-fall collapse. The heat maintaining this pressure, however, is radiated from the stellar surface, so that the star shines very brightly and shrinks slowly.

Like protostars, pre-main-sequence stars are convectively unstable. The underlying physics, however, is quite different. In general, convection in a star begins whenever the temperature drops very rapidly from the center to the surface. In protostars, deuterium burning at the center creates the convection cycle. But by the time a protostar evolves to the pre-main-sequence stage, it has exhausted its supply of deuterium.

The great luminosity of pre-main-sequence stars accounts for the steep temperature gradient within the star. The high levels of radiant energy given off cool the outer layers quickly, whereas the interior region remains insulated by the overlying matter. As the star ages and luminosity diminishes, the region of convective instability decreases. In the present sun, convection still survives over the outer 30 percent of the radius. The rising and sinking eddies create the granulated texture of the solar surface.

As the star becomes more compact, its internal temperature steadily rises and eventually reaches about 10 million kelvins. At this point, ordinary hydrogen begins to fuse into helium. The heat released from the fusion builds up the pressure to halt the contraction, and the star enters the main sequence.

It took our sun, a typical hydrogen-burning star, about 30 million years to contract from its large protostar radius to its present size. The heat released from the subsequent hydrogen fusion has maintained this size for about five billion years.

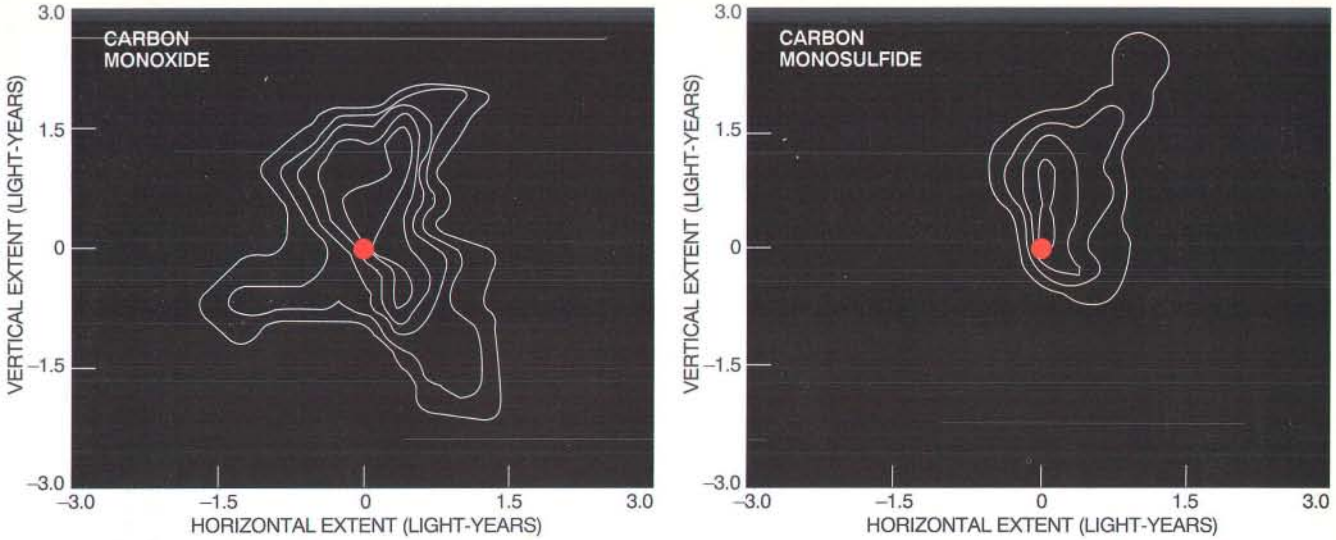
The descriptions of stellar evolution I have just given are consistent with physical theories and known nuclear processes. But theory needs to be supported by data. The data consist of measurements of the properties of many stars at different stages of their evolution. The most convenient way to express such data is to display graphically the evolution of optically visible stars in the Hertzsprung-Russell (H-R) diagram.

The H-R diagram is a graph that plots stellar luminosity on the vertical axis and surface temperature on the horizontal axis. Main-sequence stars like the sun, which fuse ordinary hydrogen, lie along a diagonal curve. Theoretical calculations show that the luminosity and surface temperature of a hydrogen-burning star—and thus its position on the curve—depend only on its mass.

This theory agrees well with observation. Astronomers determine the luminosity of a star by measuring its brightness (provided that the distance to the star is also known) and deduce the surface temperature by analyzing the star's spectrum. When one measures these two quantities for a given cluster of stars and plots the data on the H-R diagram, most of the stars indeed lie along the theoretical main-sequence curve.

Because a pre-main-sequence star is more luminous than a main-sequence one of the same mass, it will lie above the main-sequence line in the H-R diagram. The luminosity decreases with time because the shrinking of the star diminishes the surface area from which the radiation can be emitted. As a result, the star's representative point slides along a definite path that is the same for all stars of its mass. Astronomers refer to this path as a Hayashi track, in honor of Chushiro Hayashi of Kyoto University, who first calculated the properties of pre-main-sequence stars in the early 1960s.

Observations of nearby young clusters—that is, stars with a lot of interspersed molecular gas between them—have revealed that many of the stars lie above the main sequence. Those that lie near Hayashi tracks corresponding to one solar mass or less are known as T Tauri stars. Their more massive counterparts are called Herbig Ae and Be

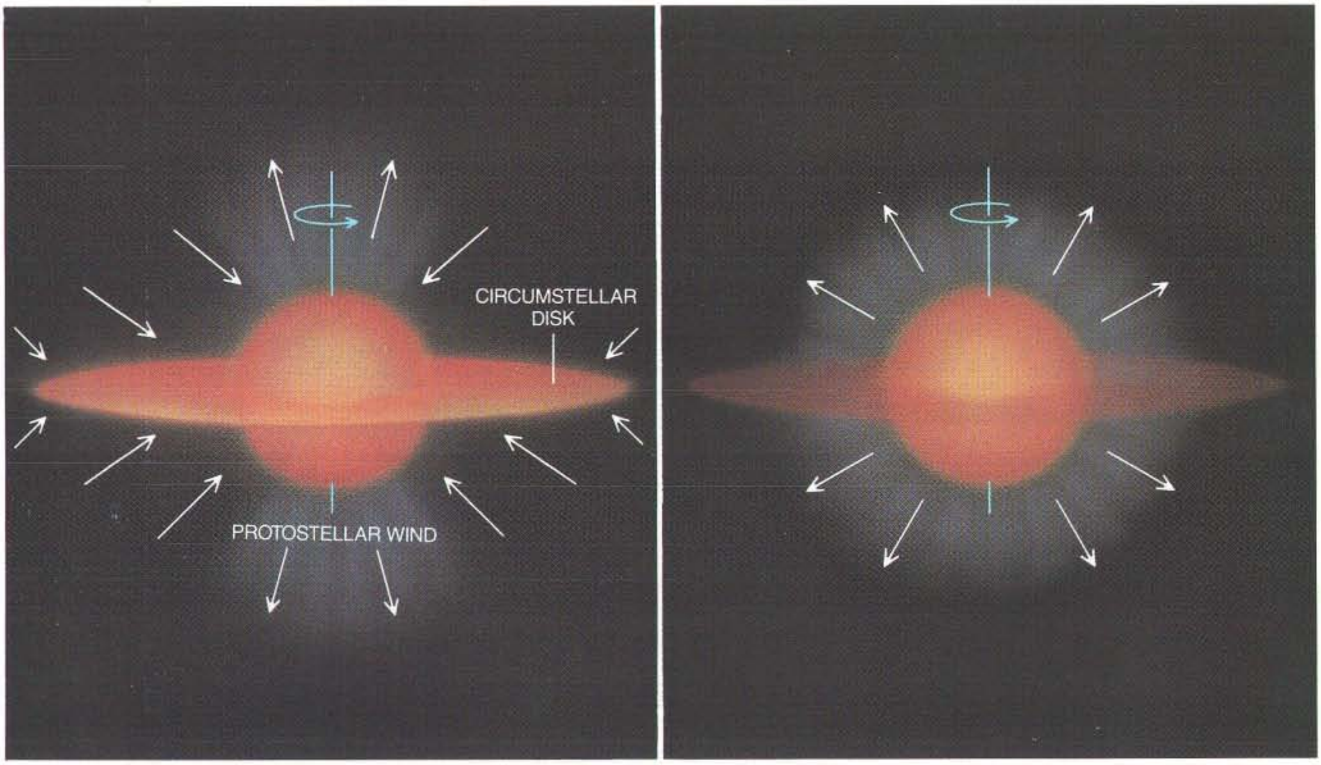


**RADIO MAPS of a dense core known as B5, about 1,000 light-years away in the Perseus constellation, are generated by analyzing the weak, millimeter-wavelength emission of carbon monoxide (*left*) and carbon monosulfide (*right*). The red dot in the middle of each image represents an infrared point source—very likely a protostar.**

stars. (The latter group was named for George Herbig of the University of Hawaii, the astronomer who pioneered the observational study of young stars.)

Although theorists are gratified that many stars lie above the main sequence, it is a more difficult matter to prove that these stars are actually descending their appropriate Hayashi tracks. Recall that deuterium burning in protostars gives them a definite radius for each value of mass. In 1983 I used this relation, together with the set of known Hayashi tracks, to make a prediction: once pre-main-sequence stars become optically visible, they should all appear along another curve in the H-R diagram. From this curve, called the birthline, each star descends along its appropriate Hayashi track to the main sequence.

Observations appear to bear out the idea of the birthline. In 1979 Martin Cohen and Leonard V. Kuhi of Berkeley published a systematic study of hundreds of T Tauri stars. In 1984 Ulrich Finkenzeller of the Landessternwarte Königstuhle (the state observatory) in Heidelberg and Reinhard Mundt of the Max Planck Institute for Astronomy presented a similar survey for the rarer



**PROTOSTELLAR WINDS are thought to be directed in a bipolar fashion along the rotation axis (*left*). The wind blows back incoming gas, but apparently does not disrupt the circumstellar disk. Eventually the star becomes visible as a pre-main-sequence object, with a much weaker wind (*right*). The cause of the protostellar wind remains a mystery.**

Herbig Ae and Be stars. The measured luminosities and surface temperatures of these stars lie on or below a well-defined boundary in the H-R diagram. The boundary coincides well with the theoretical birthline. Furthermore, those visible stars at the center of molecular gas outflows also lie along the birthline. Their location in the diagram confirms the fact that the outflow phase is associated with the beginning of pre-main-sequence contraction.

Palla and I have shown that the birthline must intersect the main sequence at some point. We calculated that the two curves meet at a position corresponding to a stellar mass of eight solar masses. In physical terms, the finding means that any star more massive than this critical value actually begins to fuse ordinary hydrogen while its parent dense core is still collapsing onto its surface. These massive stars should therefore never exhibit a visible pre-main-sequence phase. So far this prediction also seems to be in accord with existing observations.

Despite this encouraging success of the theory, many of the known properties of young stars are still not entirely understood. Most young stars, for example, are irregular variables: their luminosities fluctuate over periods ranging from hours to months. The spectra of many T Tauri stars, the group that has been particularly well monitored, show far more infrared and ultraviolet radiation than do main-sequence stars of similar mass. Yet Frederick Walter of the State University of New York at Stony Brook has found that other T Tauri stars, with very similar masses and ages, exhibit almost no excess emission. Finally, there is much evidence for strong stellar winds. These winds could well be the remnants of the much more powerful ones believed to end the protostar phase.

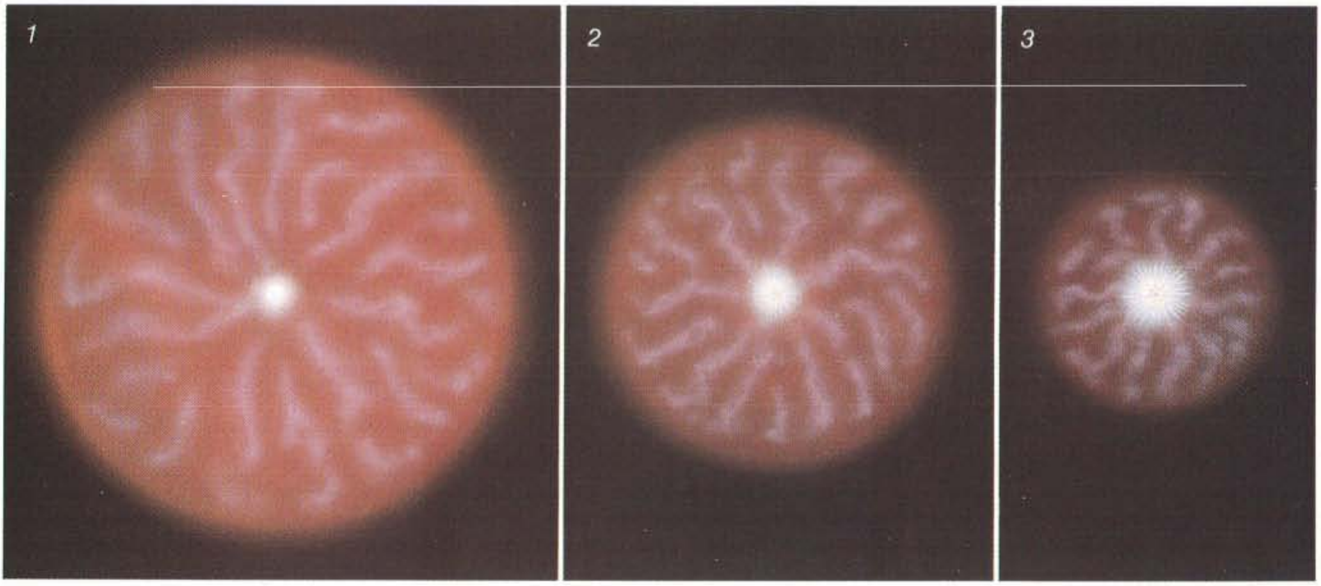
The models describing the birth of stars indicate an important byproduct: the circumstellar disk. Investigators believe such disks provide the raw materials from which planetary systems form [see "Worlds around Other Stars," by David C. Black; SCIENTIFIC AMERICAN, January]. A disk forms because not all the material collapsing inside a dense core directly joins the protostar. Whatever process formed the dense core almost certainly imparted some degree of rotation as the collapse began. Within the rotating core, the gas with the highest angular momentum lies farthest from the polar axis. As the region of collapse spreads outward, it engulfs the more distant gas. This material begins to fall inward but misses the protostar. The gas goes into orbit around the protostar, assuming the form of a circumstellar disk.

The manner in which the direction of the falling gas gradually shifts from the protostar to the disk was first worked out mathematically in 1976 by Roger Ulrich of the University of California at Los Angeles and, independently, in 1981 by Patrick Cassen and Anne Moosman of the National Aeronautics and Space Administration Ames Research Center. Cassen and Moosman also first investigated the theoretical physical properties of disks, such as their sizes and surface densities. Great interest currently exists in extending their work to the older disks surrounding pre-main-sequence stars. This focus has arisen not only because such research promises to illuminate how planets form but also because recent observational evidence indicates the existence of disks.

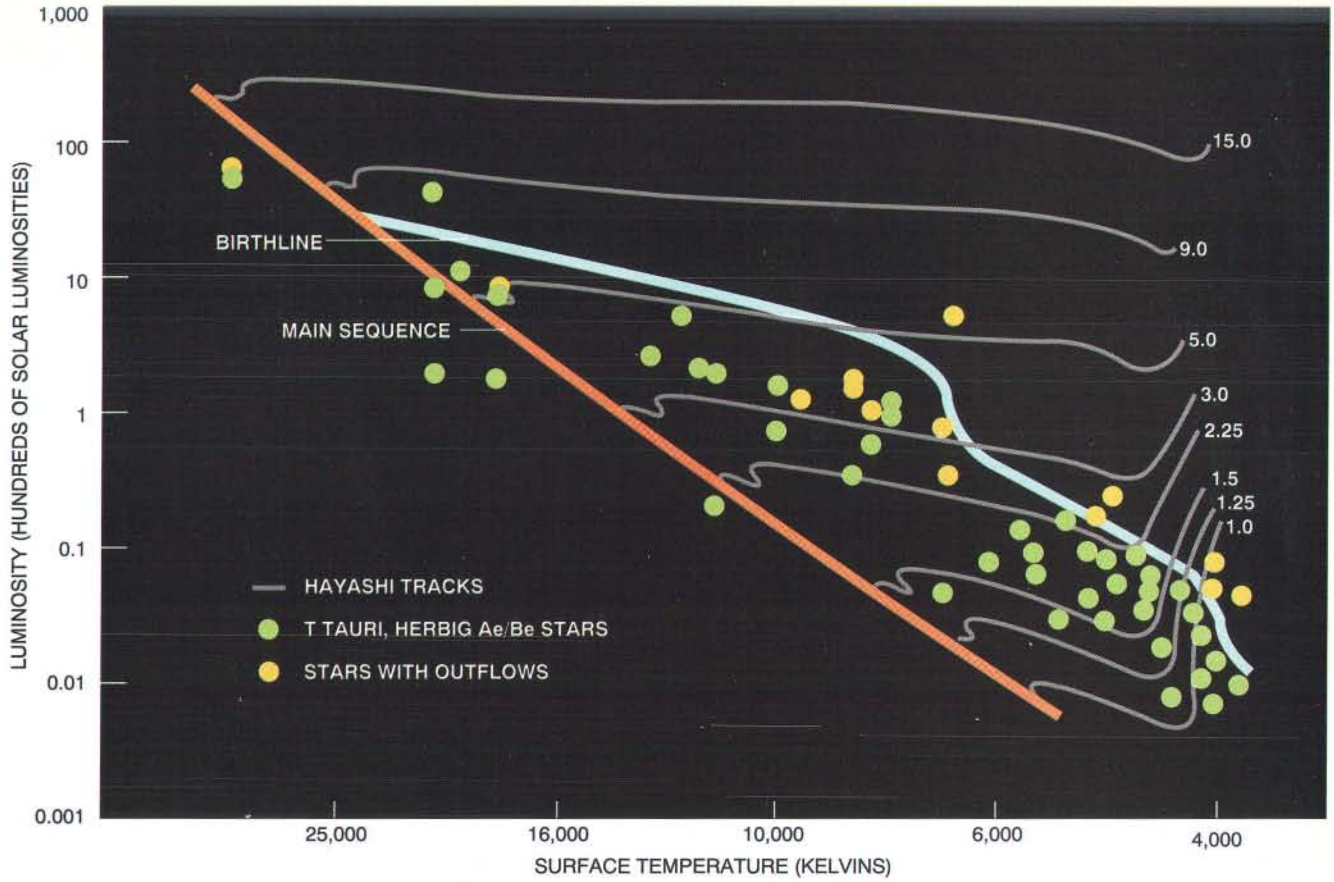
One kind of evidence consists of actual images that show circumstellar matter around young stars. In 1987, for example, Steven Beckwith of Cornell University and Anneila Sargent of the California Institute of Technology detected extended emission from carbon monoxide gas surrounding the T Tauri star HL Tau. They attributed this emission to a low-mass disk with a diameter of several light-weeks.

Another kind of evidence for disks is more indirect and hence more controversial. The evidence is actually a deduction: theorists assert that certain observed properties of T Tauri stars can best be explained by the presence of disks. Following the original suggestion in 1974 by Donald Lynden-Bell and James Pringle of the University of Cambridge, investigators have commonly attributed both the infrared and ultraviolet excesses of these stars to luminous disks that are continuously transporting mass onto their host stars.

Material must somehow lose angular momentum if it is to spiral onto the star. Lynden-Bell and Pringle assumed that an unspecified friction exists throughout the disk. As two adjacent rings of gas rub against each other,



**PRE-MAIN-SEQUENCE STAR becomes dimmer as it contracts under its own gravity. Heat loss at the surface drives convection throughout the interior, and the central temperature begins to rise. When the center reaches about 10 million kelvins, hydrogen fuses to form helium. The fusion reaction releases tremendous amounts of energy, which halts further contraction. At this point, the star is considered to have reached the main sequence and will burn steadily for billions of years.**



**HERTZSPRUNG-RUSSELL DIAGRAM plots stellar luminosity against surface temperature. Pre-main-sequence stars begin at the birthline and travel along distinct paths, called Hayashi tracks, before reaching the main sequence. The observed positions of T Tauri and Herbig Ae and Be stars and stars with outflows are consistent with the theory of star formation: they lie between the birthline and the main sequence and appear to be following their appropriate Hayashi tracks. Each track in the figure is labeled by the corresponding stellar mass in solar units. The axes scale logarithmically.**

the friction would cause the inner, more rapidly rotating ring to slow down and contract, just as the orbit of a satellite above the earth slowly decays because of atmospheric drag.

According to this picture, the excess infrared emission represents the heat generated by the friction. The ultraviolet radiation, on the other hand, is supposed to arise from a narrow, hot region between the disk and the star, in which an even stronger frictional force brakes the orbiting gas. Using models of this kind, workers have successfully matched many spectral features of T Tauri stars. They are Lee W. Hartmann and Scott Kenyon of the Harvard-Smithsonian Center for Astrophysics, Gibor Basri of Berkeley and Claude Bertout of the University of Paris.

Yet despite the best efforts of theorists over many years, no plausible explanation exists for the internal friction that these models posit. Calculations show, for example, that ordinary molecular viscosity is far too small to cause appreciable spiraling of the gas onto the central star.

In my opinion, the failure to explain the source of the friction implies that the underlying model is inadequate. A better approach may be to drop the assumption of internal friction altogether. Theorists should look again at the structure of disks that can actually form during the collapse of rotating dense cores. During the past several years, my students and I have undertaken just such an investigation. Thus far the models of disk formation we have obtained are different from the frictional ones.

All current observations of disks, both direct and indirect, indicate masses that are merely a fraction of the central star's mass, perhaps a few percent or less. Theorists find this fact disturbing and challenging. For if the accumulation of collapsing material with excess rotation builds up disks, why should the process stop so soon after the star itself was formed? If indeed protostellar winds halt the collapse phase, do the small disk masses indicate a causal link between disk formation and the triggering of these energetic outflows?

Such questions still have no answers. But these unsolved problems should be viewed as gaps in a chain whose main links have already been forged, through an extraordinary interplay of experimental and theoretical work. We can close these gaps and complete the story of young stars if we can but read the clues nature has given. And those clues are right over our heads, winking in the clear night sky.

### FURTHER READING

STAR FORMATION IN MOLECULAR CLOUDS: OBSERVATION AND THEORY. Frank H. Shu, Fred C. Adams and Susana Lizano in *Annual Review of Astronomy and Astrophysics*, Vol. 25, pages 23–81; 1987.

IN DARKNESS BORN: THE STORY OF STAR FORMATION. Martin Cohen. Cambridge University Press, 1988.

UNDERSTANDING YOUNG STARS: A HISTORY. Steven W. Stahler in *Publications of the Astronomical Society of the Pacific*, Vol. 100, No. 634, pages 1474–1485; December 1988.

THE BIRTHLINE FOR INTERMEDIATE-MASS STARS. Francesco Palla and Steven W. Stahler in *Astrophysical Journal*, Vol. 360, No. 2, Part 2, pages L47–L50; September 10, 1990.

# Stroke Therapy

*Although no treatment to limit brain damage is yet available, many tantalizing possibilities are on the horizon. Some are already being evaluated in all-important human trials*

by Justin A. Zivin and Dennis W. Choi

If you suffered a stroke today—localized brain damage caused by diminished blood flow—physicians would be unable to limit the extent of damage, no matter how quickly they reached you. At most, they would treat any medical complications, keep you comfortable and try to prevent a recurrence. This unfortunate scenario may soon change, now that researchers in many laboratories are developing ways to reduce brain injury associated with stroke. Indeed, some therapies are already being tested in human trials.

The need for interventions is great, both because of the seriousness of the disorder and its prevalence. By attacking the brain, stroke hits at what makes us human, killing or disabling an apparently healthy person in minutes. Furthermore, it ranks among the top killers in the industrialized nations.

In the U.S., according to the National Institutes of Health (NIH), the condition is the third leading cause of death and the most common cause of adult disability. Of the approximately 500,000 new victims each year, roughly 30 percent die, and 20 to 30 percent become severely and permanently disabled. Some are left completely unconscious. Others suffer paralysis, impaired cognition, reduced coordination, visual disturbance, loss of sensation or some combination of effects. Physical therapy can help many people make the best of their remaining capabilities, but it cannot repair the brain damage itself.

JUSTIN A. ZIVIN and DENNIS W. CHOI are practicing physicians as well as research scientists. Zivin is professor of neurosciences at the University of California, San Diego, School of Medicine and a staff neurologist at the San Diego Veterans Administration Medical Center. Before moving to California, he spent seven years in the departments of neurology and pharmacology at the University of Massachusetts Medical Center in Worcester. Choi recently left Stanford University to become Andrew B. and Gretchen P. Jones Professor and head of neurology at the Washington University School of Medicine in St. Louis. He is also neurologist in chief at Barnes Hospital in the same city.

At one time neurologists despaired of ever being able to limit stroke-related tissue destruction. That pessimism has lifted in the past several years, as investigators have begun to understand better the normal functioning of the brain and to unravel the details of how ischemic, or blood-deprived, brain cells die.

The brain depends on blood for a continuous supply of oxygen and glucose. Interrupt the blood flow for only a few minutes, and certain highly vulnerable neurons will degenerate; sustain the interruption, and all types of brain cells in a blood-deprived area will die, including both neurons and supportive cells known as glia. Usually tissue damage is irreversible within hours after a stroke begins.

It would be reasonable to guess that the affected cells die directly from a loss of energy. Yet evidence collected more than 15 years ago by several investigators, including Konstantin A. Hossmann of the Max Planck Institute for Neurological Research in Cologne and Bo K. Siesjö of the University of Lund, suggested that the steps linking ischemia to neuronal death were considerably more complicated than that. Details of the destructive process have now begun to emerge. It seems that some ischemic neurons are probably killed by damaging cascades of chemical interactions, cascades that begin when ischemia perturbs normal brain-signaling processes.

Current optimism that treatments might be developed stems partly from the hope that a number of brain cells might be spared by drugs that block initiation or progression of these destructive cascades. Investigators are additionally studying the value of promptly restoring blood flow to the ischemic brain, such as by removing an obstructive blood clot. Most strokes are caused by clots that either form at the site of occlusion in a cerebral artery or travel there from the heart.

The clot-removal strategy was the first to be tested, although the earliest experiments, carried out some 20 to 30 years ago, did not augur well for the approach. Neither surgery nor administration of clot-dissolving enzymes—at the time, either streptokinase or urokinase—improved the survival rate or functioning of patients.

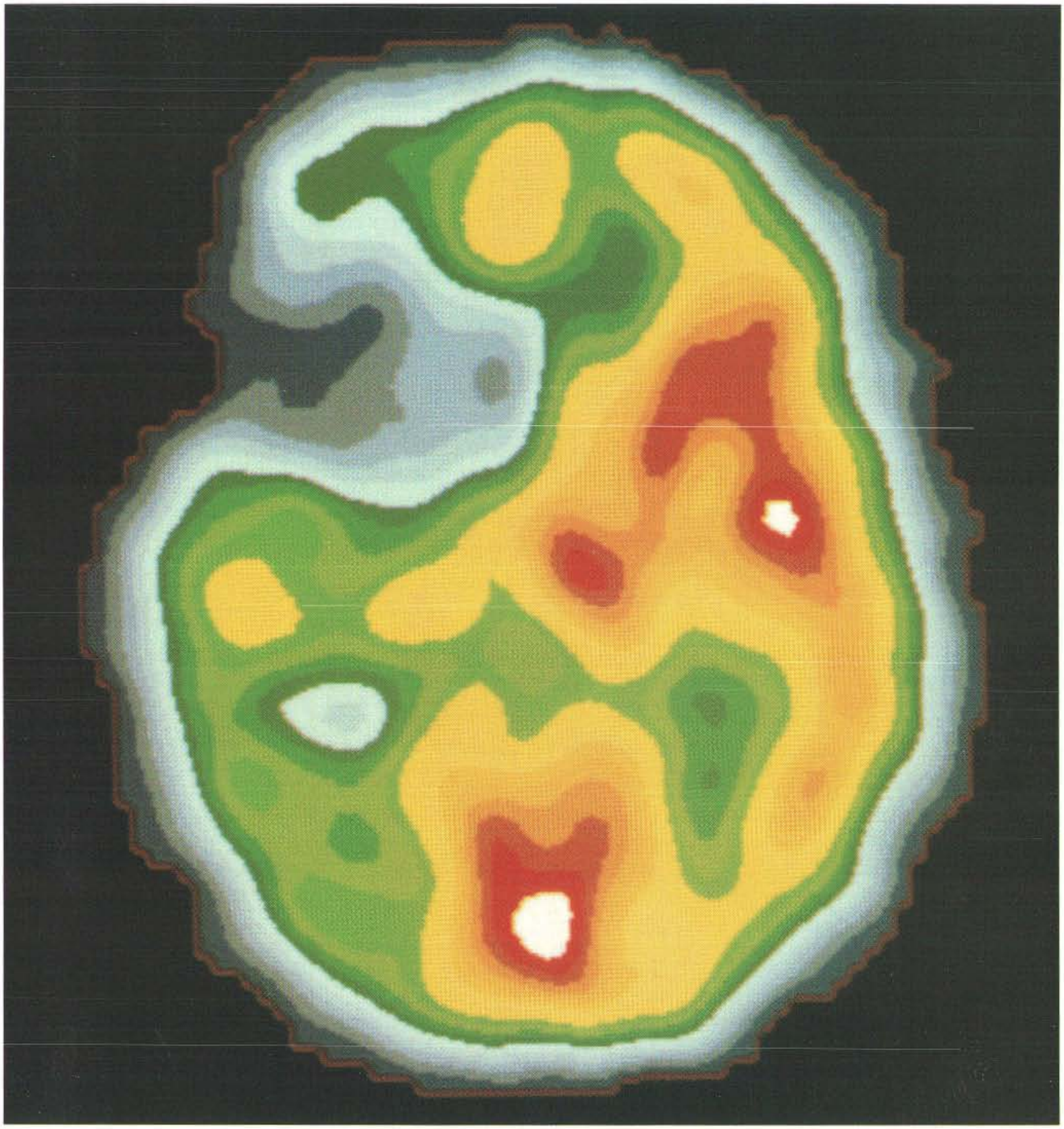
To some extent, the inability to demonstrate improvement may have derived from poor timing: more than eight hours usually elapsed between obstruction and clot removal. During that span, many cells probably died from the ischemia itself. Moreover, dangerous hemorrhaging often followed the delayed restoration of blood flow. The risk of hemorrhage increases with time because blood vessels in ischemic areas can become damaged and prone to rupture. (Today, as then, surgery can rarely be completed promptly enough to be helpful.)

The inability of workers to identify, and eliminate from consideration, patients who were hemorrhaging from the start might have contributed to the poor results as well. Some strokes are caused not by a clot but by rupture of a blood vessel, leading to a combination of mechanical injury and ischemic injury. In these cases, clearing a clot would not ameliorate the hemorrhagic damage. It is possible, too, that infusion of streptokinase or urokinase, each of which remains active in the blood for hours, actually induced some bleeding in the subjects by impairing normal blood clotting.

Fortunately, the disappointing early findings did not eliminate research into the clot-removal strategy. Today another compound, tissue plasminogen activator (tPA), is renewing excitement. A chemical made by the body as well as in the laboratory, it normally removes blood clots when they are no longer needed, such as after a wound has begun to heal.

About 10 years ago Désiré Collen and his colleagues at the University of Leuven in Belgium found a way to produce tPA in the quantities needed for small studies of its therapeutic value. They and others then discovered that the product remained active in the bloodstream for only about 10 minutes, which suggested that tPA should be less likely than either streptokinase or urokinase to induce hemorrhage. It might therefore prove to be safer for stroke victims.

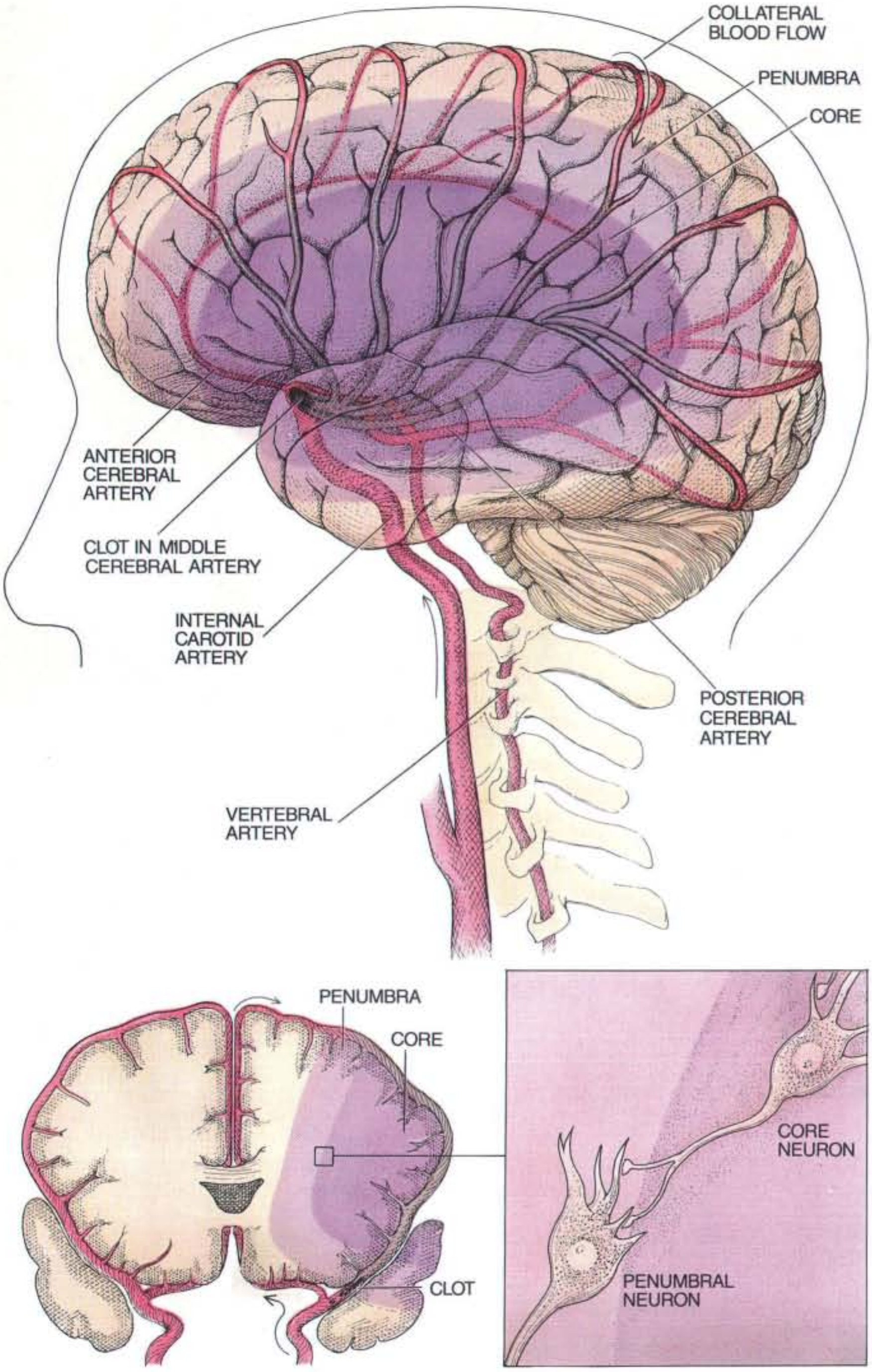
Several investigators at different institutions then established that infused tPA rapidly dissolves obstructions in the cerebral arteries of animals. One of us (Zivin) confirmed these reports and demonstrated that tPA limited tissue damage and diminished functional loss in animals treated within an hour after



**BRAIN OF A STROKE VICTIM was imaged by positron emission tomography; the front-to-back section here highlights the deficit in blood flow (ischemia) responsible for the man's brain damage. The front of the brain is at the top of the picture; the left hemisphere is at the left. The rate of flow is lowest in the injured left frontal region, where gray and bluish hues are most prominent. The moderate flow in the rest of the hemisphere (indicated mainly by greens) is below normal, but the tissue survives. The right hemisphere is fully perfused. Had treatments been available, the area of tissue destruction might have been smaller. William J. Powers of the Jewish Hospital in St. Louis provided the photograph.**

a stroke began. Indeed, animals that would otherwise be expected to die not only survived but behaved virtually normally after therapy. Zivin's group showed as well that administration of tPA need not increase the risk of brain hemorrhage.

The real test of any drug is, of course, its effect in humans, but clinical trials had to await mass production of the compound, a feat achieved about seven years ago by Genentech, Inc. Two major safety trials, one sponsored by NIH using Genentech's tPA and the other by the Burroughs Wellcome Company, ensued.



**TISSUE AT RISK** of harm from occlusion of a cerebral artery can be divided into two regions: the core (containing cells that are highly dependent on the blocked artery) and the penumbra (containing surrounding cells that receive some blood from other arteries). Core cells are probably overwhelmed by many destructive processes. Penumbral cells may be damaged most by events begun when ischemia induces neurons to oversecrete glutamate, an excitatory neurotransmitter.

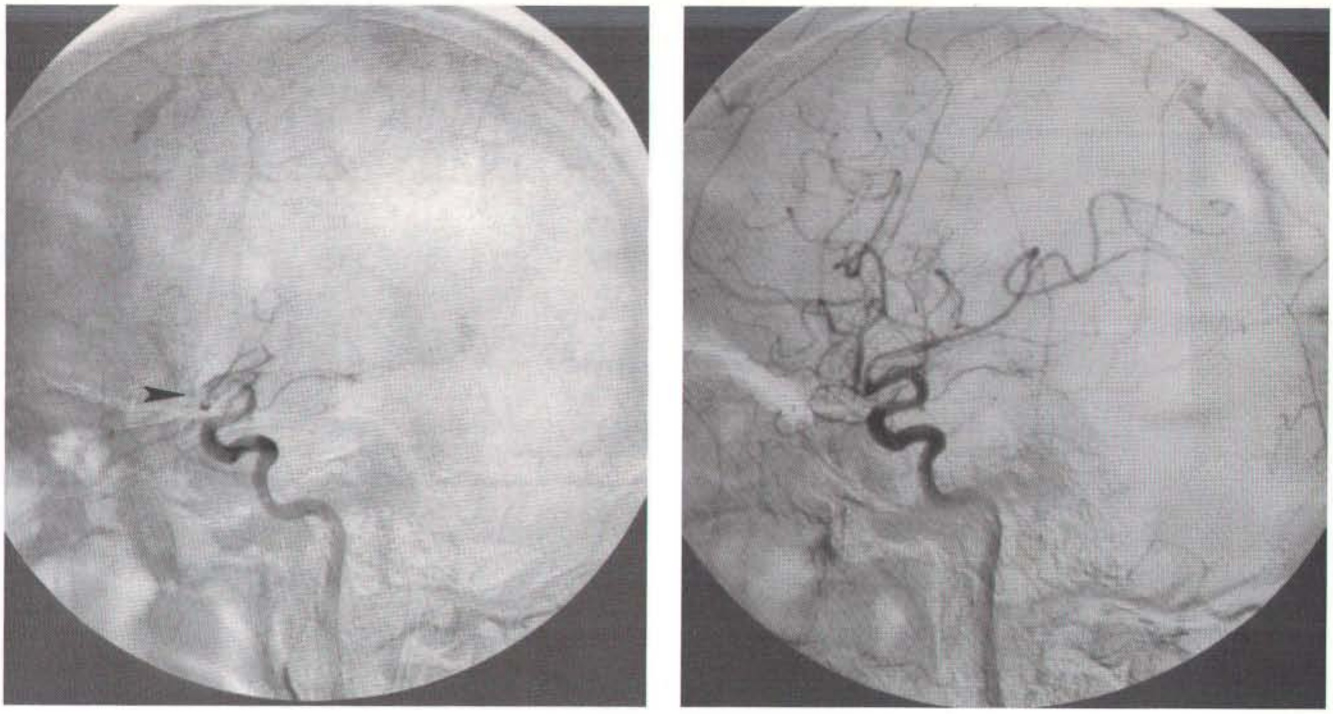
The studies examined a combined total of more than 170 patients, who were treated either within one and a half or eight hours after the onset of stroke, respectively. Both projects found that the drug rarely causes severe complications at the doses tested. They also obtained some apparent evidence of effectiveness. For example, the Burroughs team found that in approximately a third of its patients, the cerebral blockage cleared partly or fully within an hour after treatment. A second NIH study, designed to evaluate directly the drug's ability to improve functioning in stroke patients, is under way at many hospitals. Results are expected in early 1994 or before.

Researchers are also intensively studying ways to halt the chemical events leading from ischemia to brain damage. A number of the experimental therapies attack what may be called the glutamate cascade, a sequence we and many other investigators think is a major cause of ischemia-induced neuronal death. Cells at the center of the stroke—in the "core" region that receives essentially no blood when a vessel is occluded—die from several overwhelming causes and probably cannot be salvaged by any treatment short of immediate clot removal. In contrast, the glutamate cascade could well be a leading cause of damage in the surrounding, "penumbral," region, which receives some blood from nonoccluded vessels.

The cascade begins after terminals of ischemic neurons release excessive amounts of the neurotransmitter glutamate, an excitatory stimulus, into the spaces between cells. (Many, perhaps a majority, of the neurons in the brain contain this transmitter.) The cells are initially prompted to secrete glutamate by depolarization of the outer membrane. Normally, energy-requiring pumps force ions into and out of the cell in an attempt to ensure that the membrane remains polarized. When the cells lose energy, the pumps fail.

In healthy tissue, neurons and glial cells would remove excess glutamate from extracellular spaces, but ischemic cells lack the energy they need for the job. Consequently, when neurons release large amounts of glutamate, the neurotransmitter binds in quantity to receptor molecules on neighboring neurons. This sustained binding induces an abnormal movement of calcium ions ($Ca^{2+}$) into the recipient cells, producing a buildup of calcium, which apparently contributes to many of the events that destroy those cells.

The idea that neurotransmitter activ-



**OBSTRUCTION of the middle cerebral artery (*arrow at left*), which feeds the frontal, temporal and parietal lobes, was cleared in a patient (*right*) by a compound called tissue plasminogen activator (tPA). The person's ability to move improved rapidly. Prompt removal of blockages by this or other means is probably the only way to rescue cells of the core.**

ity could be a critical link between ischemia and neuronal death first gained credence in the early 1980s, when Ira S. Kass and Peter Lipton of the University of Wisconsin at Madison and Steven M. Rothman of Washington University in St. Louis did seminal experiments on neurons of the hippocampus, a brain region particularly vulnerable to hypoxia (oxygen deprivation).

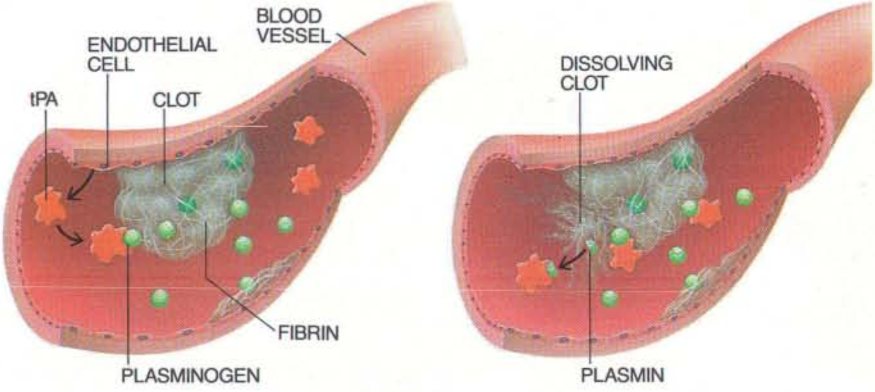
Knowing that magnesium impairs the ability of neurons to release neurotransmitters, Kass and Lipton, working with brain slices, and Rothman, working with cells in culture, added extra magnesium to the bathing solutions and then deprived the slices or cells of oxygen. The magnesium reduced the expected damage, implying that inhibition of signaling can protect neurons from hypoxia and, hence, that some neurotransmitter might play a part in injuring oxygen-deprived brain cells during a stroke.

The question was, Which neurotransmitter? Convergent evidence suggested the culprit could be glutamate. Work by John W. Olney and his collaborators at Washington University in the 1970s first established that excessive exposure to glutamate or related compounds can kill brain neurons, presumably by overstimulating them. Olney named the killing process excitotoxicity.

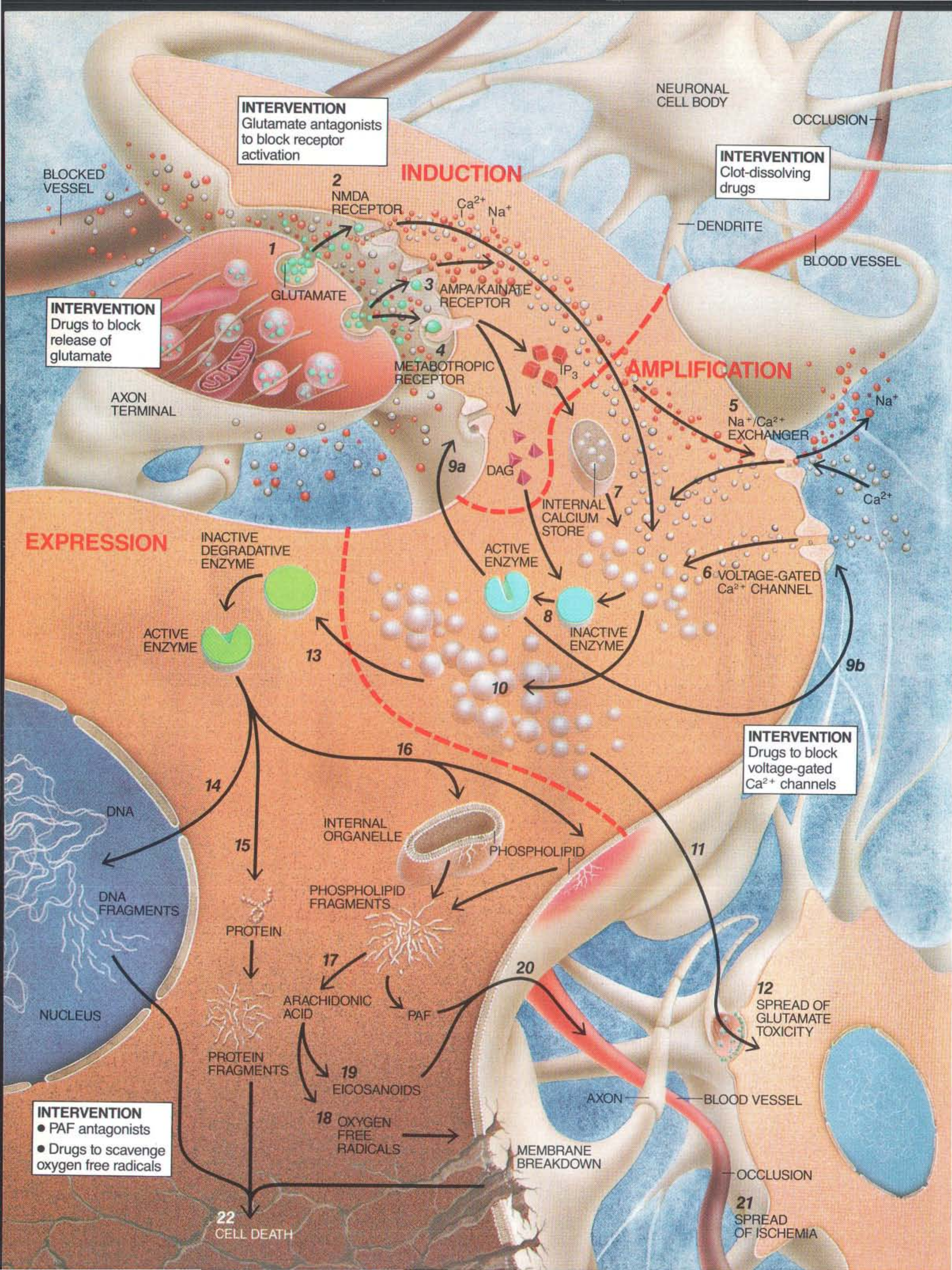
Subsequently, Helene Benveniste, Nils H. Diemer and others at the University of Copenhagen showed that when brain tissue is made ischemic, the levels of glutamate outside the affected brain cells rise. At about the same time, Rothman, studying hippocampal neurons in cell culture, and Roger P. Simon, Brian S. Meldrum and others at the Institute of Psychiatry in London, while studying hippocampal cells in rats, demonstrated that glutamate antagonists (which block activation of glutamate receptors) can decrease the loss of hypoxic neurons.

The work of these investigators did more than support a primary role for glutamate; it also raised a key question about how the neurotransmitter exerts its toxic effects. Glutamate excites neurons by stimulating several subtypes of glutamate receptors. Yet the antagonist Simon and his colleagues tested, 2-amino-7-phosphonoheptanoate (then newly developed by Jeff C. Watkins and his colleagues at the University of Bristol), was more specific: it selectively blocked only one subtype, that known as the *N*-methyl-D-aspartate (NMDA) receptor. How could such selective blockade successfully protect neurons?

Studies conducted by one of us (Choi) provided a possible answer to the ques-



**tPA DISSOLVES BLOOD CLOTS by activating the enzyme plasmin. When a clot forms, plasminogen, a precursor of plasmin, becomes trapped among the fibrin strands that constitute the bulk of the clot (*left*). tPA, which is naturally made by endothelial cells, cleaves plasminogen on the surface of the clot (*right*), thereby releasing active plasmin. The plasmin, in turn, begins to degrade fibrin and thus to expose more plasminogen. The process continues until the clot is gone.**



**INTERVENTION**
Glutamate antagonists to block receptor activation

**INTERVENTION**
Clot-dissolving drugs

**INTERVENTION**
Drugs to block release of glutamate

**INTERVENTION**
Drugs to block voltage-gated $Ca^{2+}$ channels

**INTERVENTION**
- PAF antagonists
- Drugs to scavenge oxygen free radicals

## The Stages of Stroke

Excess release of glutamate may damage brain tissue by a three-stage process. Promising strategies for impeding each proposed stage are under investigation (*boxes*).

**STAGE 1:** INDUCTION
After a blood vessel becomes occluded (*darkened vessel at left and right*), neurons (*such as the cell whose axon terminal is shown at the top left*) become starved for blood and, thus, for oxygen and glucose. The blood-deprived neurons then release excessive glutamate (***1***), which activates glutamate receptors on other neurons (*such as on the large neuron shown in the cutaway*). In response to glutamate, NMDA receptors (*top center*) open channels that allow passage of both sodium ($Na^+$) and calcium ($Ca^{2+}$) (***2***), and AMPA/ kainate receptors open $Na^+$ channels (***3***), resulting in an abnormal buildup of both ions. At the same time, metabotropic receptors trigger formation of inositol 1,4,5-trisphosphate ($IP_3$) and diacylglycerol (DAG) (***4***).

**STAGE 2:** AMPLIFICATION
Excess $Na^+$ activates a transporter that exchanges $Na^+$ for $Ca^{2+}$ (***5***); internal positive charges activate voltage-gated $Ca^{2+}$ channels (***6***); and $IP_3$ releases $Ca^{2+}$ from intracellular stores (***7***). The resulting $Ca^{2+}$ overload, combined with excess DAG, activates enzymes (***8***) that increase cellular sensitivity to glutamate and other excitatory stimuli (***9a***) and increase the contribution of the voltage-gated $Ca^{2+}$ channels (***9b***). Further buildup of $Ca^{2+}$ (***10***) triggers the release of glutamate (***11***), thus spreading the toxic cascade to other cells (***12***).

**STAGE 3:** EXPRESSION
$Ca^{2+}$ activates enzymes (***13***) that degrade DNA (***14***), proteins (***15***) and phospholipids (***16***). Phospholipid breakdown leads to the formation of arachidonic acid (***17***), which, when metabolized, gives rise to oxygen free radicals (***18***) that harm the cell membrane and to eicosanoid molecules (***19***). Eicosanoids, together with activated platelet-activating factor (PAF), another by-product of phospholipid breakdown, promote blockage of previously healthy vessels (***20***) and the spread of ischemia (***21***). Meanwhile continuing destruction of cellular constituents kills the neurons (***22***).

tion by suggesting that activation of NMDA receptors may be required for glutamate-induced killing of cells. Such a crucial role for the NMDA receptor would be consistent with Choi's demonstration that glutamate-induced death of cultured neurons from the cerebral cortex could be reduced by removal of calcium from the extracellular bathing fluid. That observation had suggested that calcium entry into the cell was a major mediator of glutamate's destructive effect. The findings implicating NMDA receptors and calcium in cell death fit with experiments by Amy B. MacDermott and her collaborators at the NIH showing that NMDA receptors are the only glutamate receptors able to open channels that allow large amounts of calcium to pass into neurons across the cell membrane.

The steps connecting glutamate overexposure and calcium accumulation to neuronal death have yet to be fully delineated. On the basis of work done in several laboratories, however, we can outline a hypothetical sequence consisting of three basic stages, each of which might be amenable to intervention.

The first stage, induction, encompasses the initial overstimulation of glutamate receptors and the immediate intracellular effects. Glutamate-bound NMDA receptors open calcium channels, which then usher a great deal of calcium into the cell. At the same time, sodium ions ($Na^+$) rush in, both through channels that are controlled, or gated, by the NMDA receptor and through channels gated by another glutamate-receptor subtype known as the AMPA/ kainate receptor. Where sodium goes, water and chloride ions ($Cl^-$) follow, and so the cell swells. The excess sodium and calcium also disrupt, reversibly at this point, the ability of the neurons to respond normally to signals from other nerve cells.

Meanwhile a third kind of glutamate receptor, the so-called metabotropic receptor, becomes activated. This receptor does not control ion channels, but it does trigger an increase in production of two intracellular messengers, diacylglycerol (DAG) and inositol 1,4,5-trisphosphate ($IP_3$). These substances may be important in the next stage, amplification. In this phase, calcium levels soar, partly because $IP_3$ frees calcium from intracellular storage sites. Additional calcium ions flood in from the outside, transported by voltage-gated calcium channels (opened by the altered charge distribution across the cell membrane) and by a carrier molecule that exchanges sodium for calcium.

This calcium overload probably combines with elevated levels of the DAG messenger to alter the activity of several families of enzymes. These enzymes modify membrane proteins (likely including glutamate receptors), increasing the sensitivity of the neurons to excitatory signals. Such enhanced excitability may promote further accumulation of calcium and increase the release of glutamate from nerve terminals. Thus, the events of this stage may not only amplify the buildup of calcium but also contribute to a worsening of excitotoxicity in adjacent neurons.

The second stage ultimately gives way to the expression stage, during which irreversible damage occurs. Early in this phase, calcium activates enzymes that attack several classes of molecules, including nucleic acids, proteins and lipids (fats). The breakdown of phospholipids in the outer cell membrane and in membranes of internal organelles may be particularly devastating.

When phospholipids are degraded, one by-product is arachidonic acid, the metabolism of which leads to formation of highly reactive compounds called oxygen free radicals. These compounds, in turn, can initiate a chain reaction, called lipid peroxidation, capable of destroying the cell membrane.

Arachidonic acid can harm cells in other ways as well. It participates in the formation of substances called eicosanoids, which increase the aggregation of blood cells and the constriction of blood vessels. Such changes are probably aggravated by the formation of platelet-activating factor, yet another consequence of phospholipid breakdown. Together eicosanoids and platelet-activating factor can thus generate a vicious circle in which ischemic injury leads to further ischemia.

Stroke-related brain damage can potentially be limited by blocking steps in the glutamate cascade. The cascade might be halted at its inception by limiting the output of glutamate or, at least in theory, by clearing the neurotransmitter from extracellular spaces. One promising approach is the delivery of drugs that stimulate receptors of the chemical adenosine on neuronal terminals. Such stimulation inhibits the release of most neurotransmitters, including glutamate.

Alternatively, it may be possible to decrease glutamate output by interfering with the synthesis of glutamate molecules that are normally stored in the terminals. Raymond A. Swanson and others at the University of California at San Francisco have found that inhibition of synthesis with methionine sulfoximine reduces ischemic damage in rat brains. A third strategy is to lower brain temperature. This maneuver is known to diminish experimental ischemic brain damage, and Myron D. Ginsberg and others at the University of Miami have recently linked the effect to blockade of glutamate release.

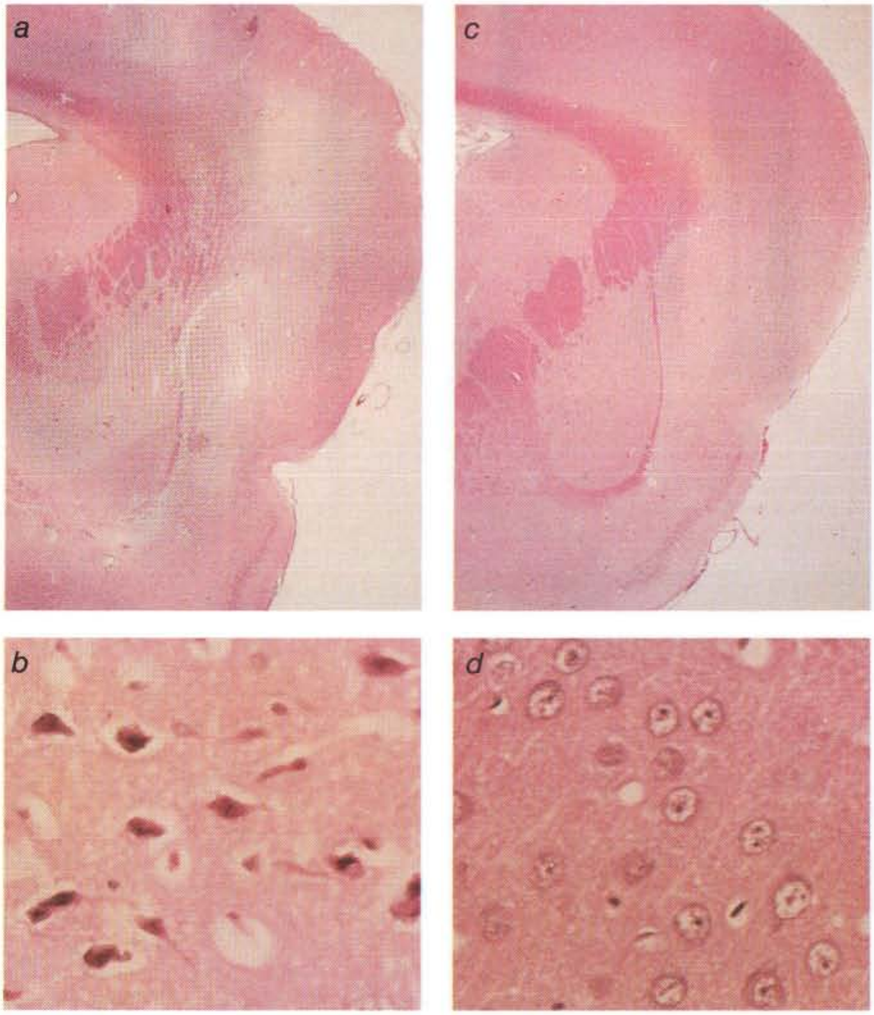
The easiest way to halt the cascade, however, might be to administer glutamate antagonists to prevent activation of NMDA receptors. Pharmaceutical companies have already developed compounds for that purpose—such as CGS 19755 (CIBA-GEIGY Corporation), MK-801 (Merck & Company) and dextrorphan (Hoffmann-La Roche, Inc.). All these drugs decrease stroke-induced tissue damage in animals; CGS 19755 and dextrorphan are now being tested for safety in humans. Furthermore, glutamate antagonists that block activation of AMPA/kainate receptors might augment the benefits of impeding NMDA receptors.

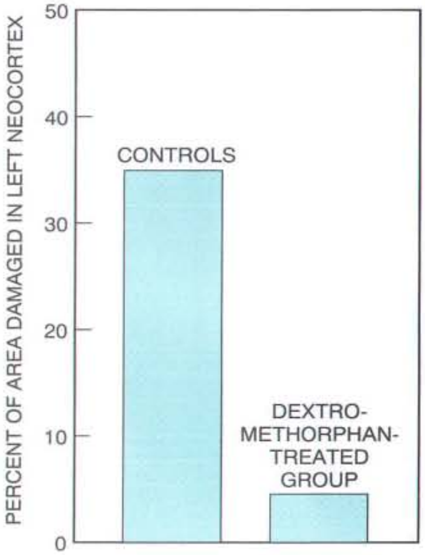
The recent development of this latter group of drugs may prove particularly important for victims of global, or brain-wide, ischemia, such as that caused when a patient bleeds profusely or when the heart stops pumping temporarily during cardiac arrest. Antagonists that act on NMDA receptors have exhibited little ability to reduce tissue damage in animals with global ischemia, whereas a new AMPA/kainate antagonist called NBQX can markedly ameliorate tissue injury in rats, according to Malcolm J. Sheardown and others at A/S Ferrosan in Soeborg, Denmark.

The reason NMDA-receptor blockers are less helpful in global than in focal ischemia may relate to the greater acidity of the extracellular fluids in the first condition. When the blood supply to the entire brain falls, the acidity of these fluids increases more than it does in the penumbra of individuals suffering from localized ischemia. Martin Morad and others at the University of Pennsylvania have found that increased acidity reduces NMDA receptor–mediated activity, but it has less effect on the damaging activities of other glutamate receptors. Drugs may be needed to accomplish that job.

Workers have also focused attention on amplification. For this stage, chemicals of the dihydropyridine class, such as nimodipine, have theoretical appeal because they pack a dual punch: they reduce the influx of calcium through voltage-gated calcium channels, and they dilate blood vessels, an effect that could improve blood flow in penumbral tissue. Dihydropyridines have yielded variable results when administered alone in clinical trials, but they may produce an additive benefit when given together with glutamate antagonists.

Protein kinase C, one of the enzymes that may participate in amplification, has been another object of investigation, although the findings are so far inconclusive. On one hand, Erminio Costa and others at the Fidia-Georgetown Institute for the Neurosciences in Washington, D.C., have found that ganglioside compounds, which are thought to inhibit the enzyme's function, can reduce ischemia-related brain damage. Similarly, Kyuya Kogure and others at Tohoku University in Sendai, Japan, have reported improved function in gerbils given other inhibitors of protein kinase C. Yet one of us (Zivin) has





**BRAIN TISSUE in an untreated rabbit was injured extensively by an experimentally induced stroke (*a and b*), but tissue in a rabbit given dextromethorphan, a glutamate antagonist related to dextrorphan, before the stroke began was largely spared (*c and d*). The damage in the first animal is evident in the broad areas of pale tissue (*a*) and in the shrunken shape of the neurons (*dark spots in b*), made visible by high-resolution microscopy. Another study has shown that dextromethorphan given an hour after a stroke begins can also limit brain damage in rabbits (*graph*). Gary K. Steinberg of Stanford University provided the micrographs and data.**

found that inhibition actually worsens neurological damage in test animals.

For patients who have reached the expression stage, interference with the activity of oxygen free radicals holds special appeal, both because the compounds are highly destructive and because their suppression probably would not interfere with the functioning of healthy cells. Possible therapeutic agents include a new class of free-radical inhibitors, 21-aminosteroids, devised by John M. McCall, Edward D. Hall and others at the Upjohn Company.

Our focus on glutamate should not imply that interference with the glutamate cascade is the only way to limit brain destruction. Several lines of research suggest that augmenting or antagonizing certain other neurotransmitters would also help protect neurons from ischemia-induced injury. In addition, studies specifically addressing how glial cells die might provide still more avenues for treatment. The events leading to their demise have yet to be elucidated fully.

Certain therapies may prove helpful even after tissue damage is complete. The first days after the onset of a stroke may be a time of critical regrowth, when somewhat damaged neurons start to regenerate injured parts, and other neurons form new connections to compensate for ones that have been lost.

One indication that treatment in this period could be fruitful comes from Dennis M. Feeney and other investigators at the University of New Mexico. The researchers suspected that a depression of the catecholamine neurotransmitter system might contribute to behavioral deficits after a stroke. When they gave amphetamines, which increase catecholamine transmission, to rats who had suffered damage to their frontal cortex, the compounds produced a lasting improvement in the animals' ability to walk across a narrow beam. Similar improvements were also seen in treated cats. Someday, even later treatment may become possible, if neuronal transplantation enables physicians to restore lost cells to the brain.

Ultimately, care givers will probably combine therapies, making choices depending on the individual needs of a patient. For instance, a glutamate antagonist may prove safe for almost anyone and might be given by an emergency medical technician in the field. But administration of tPA or similar drugs might have to be delayed until hemorrhage was ruled out with a brain scan. Patients who could safely be treated with both approaches should reap an added benefit: one of us (Zivin) has shown in animals that the combination of a glutamate antagonist and tPA reduces neurological injury more than would either drug alone. Clot-dissolving therapy may improve the ability of other drugs to reach ischemic tissue.

Although therapy remains a future goal, methods for preventing stroke are already very much a feature of the present. The incidence of the disorder has been declining in the U.S. for more than two decades—an improvement that correlates with the widespread availability of treatments for hypertension. The risk of stroke might also be lessened by controlling atherosclerosis, heart disease, diabetes and smoking, although the evidence is less compelling than for hypertension. Recent findings further indicate that surgery to remove arterial plaques can lower the incidence of stroke in selected patients with atherosclerosis. And in patients with a history of one or more warning incidents, or transient ischemic attacks, aspirin can reduce the incidence of subsequent stroke.

Of course, none of these preventive steps helps once a vessel becomes occluded, which is why the need for acute therapy is so urgent. Unfortunately, proving efficacy in humans will not be easy. Investigators need better methods for assessing the extent of tissue damage and, more important, for judging improvements in function. Because the effects of stroke differ greatly from patient to patient, depending on such factors as the region of the brain affected and the complexity of brain functions, large numbers of patients (perhaps 1,000 per trial) will be required to test the benefits of any single therapy. This is a greater number than any medical center can study in a reasonable time frame, and so cooperative, multicenter trials will have to be done.

Successful results will bring added challenges. If some of the methods outlined in this article prove worthwhile, the emergency medical care system will have to abandon its emphasis on palliation and gear up to provide prompt therapy for stroke, just as the system did when heart attacks first became treatable. Thousands of care givers will have to be trained to apply the new therapies, and the public will need to be taught the symptoms of stroke.

These are daunting tasks. Yet accomplishing them will be a delight if the elusive prize—effective treatment for stroke—can be captured. Researchers still have much to learn, but now there is reason for optimism.

**WOODROW WILSON, 28th president of the U.S., is shown in 1921, two years after he suffered a stroke that impaired control over much of the left side of his body. He fell ill while trying to convince Congress to ratify the Treaty of Versailles and approve American participation in the League of Nations. Congress did neither. American absence from the League weakened the organization and undoubtedly contributed to its inability to prevent World War II.**

FURTHER READING

MECHANISMS OF ISCHEMIC BRAIN DAMAGE. Bo K. Siesjö in *Critical Care Medicine*, Vol. 16, No. 10, pages 954–963; October 1988.

CEREBRAL HYPOXIA: SOME NEW APPROACHES AND UNANSWERED QUESTIONS. Dennis W. Choi in *Journal of Neuroscience*, Vol. 10, No. 8, pages 2493–2501; August 1990.

CEREBROVASCULAR DISEASES: SEVENTEENTH PRINCETON CONFERENCE. Durham, N.C., March 16–18, 1990. Edited by James N. Davis, Mark J. Alberts and Larry B. Goldstein. Supplement to *Stroke*, Vol. 21, No. 11; November 1990.

THROMBOLYTIC THERAPY FOR STROKE. Justin A. Zivin in *Protection of the Brain from Ischemia.* Edited by Philip R. Weinstein and Alan I. Faden. Williams & Wilkins, 1990.

# Optical Interferometry of Surfaces

*By exploiting the wave nature of light and the power of modern computers, the authors have designed highly sensitive devices for measuring surface texture*

by Glen M. Robinson, David M. Perry and Richard W. Peterson

Videotape, ball bearings, photographic film and computer chips all perform poorly if their surfaces are rough or improperly shaped. Creating precise finishes on such objects requires having precise ways to measure the texture of surfaces. Yet, until recently, the techniques available suffered from serious drawbacks.

In 1980, we set out to find a better method for examining surfaces. Optical interferometry, which exploits the wave nature of light to produce extremely accurate measurements of shapes or distances, seemed a likely candidate. Interferometry offers exemplary resolution and involves no physical contact with the surface to be studied. Unfortunately, interferometric images, or interferograms, are notoriously difficult to interpret and translate into useful measurements of surface texture. Traditional interferograms consist of snapshots of patterns of light and dark bands that often bear little obvious resemblance to the contours of the object being studied.

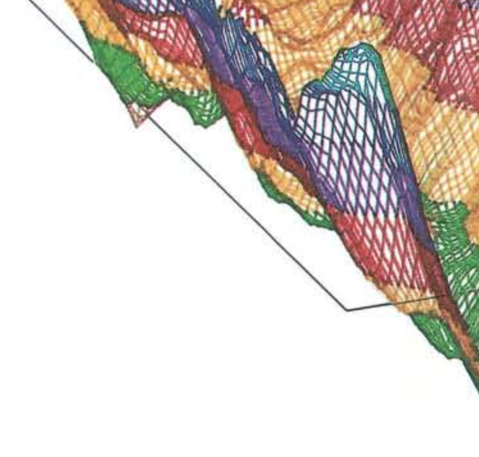
A solution to this problem occurred to one of us (Robinson) while watching a television commercial that showed three-dimensional, computer-generated images of the rock layers under an oil well. By combining interferometry with powerful modern computers and sophisticated graphics software, we realized we could present the enormous amount of information in an interferogram in an intuitively obvious and attractive manner. Moreover, once the data representing the interferogram are entered into a computer, we could use various mathematical techniques to derive useful statistics about the texture of the surface.

These methods have recently been incorporated into commercial devices manufactured by several companies. Interferometry is now used for spot-check quality control; it has led to reductions in cost and to improvements in performance of a number of products, including photographic film, magnetic tape and computer diskettes.

The rapidity with which industry has begun to embrace computer-analyzed optical interferometry indicates the inadequacies of previous measurement methods. Traditional gloss and light-scattering techniques, for example, reveal the overall texture of a surface by the way light bounces off it. Such an approach tells nothing about individual surface features or their cumulative size distribution. Optical and electron microscopes can resolve small details, but they generally cannot measure the height of the multiple bumps and blemishes found on surfaces.

Another standard method, known as stylus profilometry, involves dragging a stylus across a surface. The rising and dipping of the stylus record the terrain of the specimen's surface over a thin strip. The disadvantage of this technique lies in the great pressure of the stylus tip, which can compress and alter the surface of films and tape.

Optical interferometry offers significant advantages for studying surface texture. The only probing tool used is a low-intensity beam of light, and so the process is nondestructive. In principle, interferometry can resolve surface irregularities only a few angstroms high. (An angstrom is one ten-billionth of a meter, roughly the diameter of a hydrogen atom.) Moreover, the technique can quickly compile a statistically valid



GLEN M. ROBINSON, DAVID M. PERRY and RICHARD W. PETERSON have collaborated for more than a decade to perfect techniques for computer-analyzed interferometry. Robinson completed his Ph.D. in physical chemistry at Tulane University in 1970. One year earlier he had joined the 3M Company, where he is now a senior research specialist in the consumer video/audio products division. Perry received a master's degree in mathematics at the University of Minnesota in 1972. He taught mathematics at Bethel College and Seminary in St. Paul until 1982, when he too joined 3M. Currently he is a laboratory manager in the data-storage products division. Peterson earned a Ph.D. in physics at Michigan State University in 1969, followed by postdoctoral work at Los Alamos National Laboratory. At present, he is a professor of physics at Bethel College.

study of the surface, because the area viewed is large in proportion to the size of individual surface features.

The underlying principle of interferometry is that two light waves brought together interact, or interfere, with each other in much the same way that water waves do. In either case, if the crest of one wave coincides with the trough of the other, interference is destructive and the waves cancel out. If two waves or two troughs coincide, the waves reinforce each other. About 100 years ago the pioneering American physicist Albert A. Michelson developed techniques that exploited interferometry to make extremely precise measurements. The device he designed, known as a Michelson interferometer, is still used for many kinds of measurements.

The Michelson interferometer incorporates a partially mirrored surface, known as a beam splitter, to divide a beam of monochromatic light (light of one color or, equivalently, wavelength) into two beams traveling in different directions along the arms of the instrument. In our application, one beam reflects off a flat reference mirror; the other beam bounces off the surface of the specimen being studied. The two beams rejoin at the beam splitter.

Texture on the specimen's surface changes the distance traveled by the second beam. When the beams recombine, some parts of the second beam will be in phase with the first, others out of phase. The spatial relation between the two beams contains detailed information about the topography of the surface. This information shows up in the pattern of bright regions (where the waves reinforce each other) and dark ones (where they cancel) in the recombined beam.

A one-wavelength change in the path difference between the two interferometer arms results in the recombined wave's going through one bright-dark-bright cycle. As a result, one dark region, or fringe, is created whenever the round-trip path along one arm increases or decreases by one wavelength.

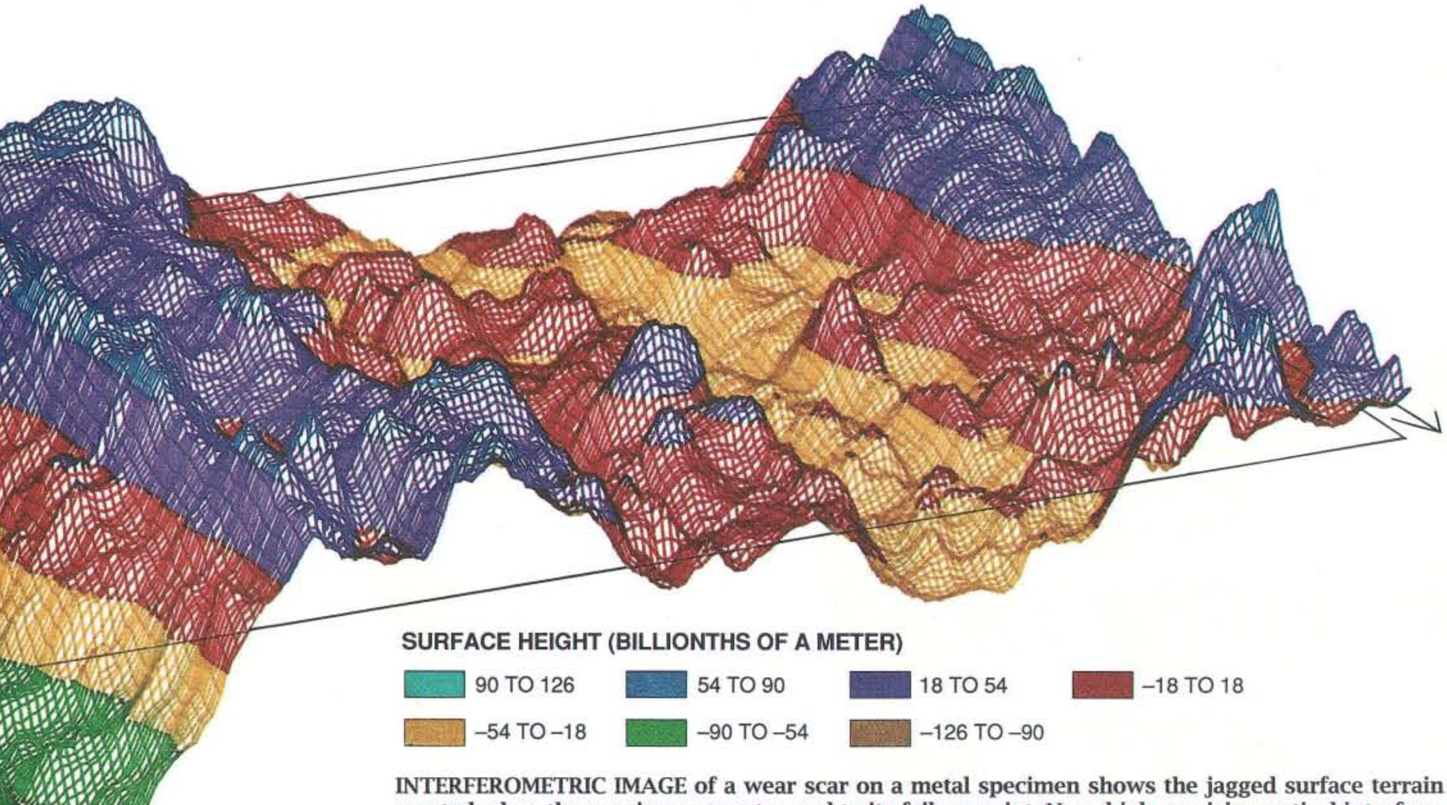
Scientists commonly use the term "phase angle" when discussing the spatial relation between two waves. A relative phase angle of 0 degrees means that the waves move in step and therefore reinforce each other. A phase angle change of 360 degrees corresponds to a path length change of exactly one wavelength along one of the arms and hence results in the creation of one additional fringe.

Tilting the reference mirror creates a continuously changing path length across the image. The result is a regular distribution of straight, parallel interference fringes. Surface features on the specimen visibly alter this pattern of fringes. If the specimen has a bump on its surface that is half a wavelength high, the fringes shift by one cycle (360 degrees), because the round-trip path grows one wavelength shorter. Fringe patterns follow the elevation of the surface much like contour lines on a topographic map. The relative spacing of the fringes is determined both by the slope of the specimen's surface and by the tilt of the reference mirror.

Manually measuring each change in the fringes of a typical interferogram can easily generate 10,000 data points. Processing that quantity of data is highly impractical without the aid of a computer. Therefore, in the past, microscopic interferometry has largely been applied to well-defined, two-dimensional problems, such as calculating the depth or height of isolated deposits or scratches.

Two advances have made possible rapid and convenient interferometric surface studies. One was the advent of relatively inexpensive high-speed and high-capacity digital computers. In 1974 John H. Bruning of AT&T Bell Laboratories provided the other key development, a process he called direct phase-detecting interferometry.

Bruning's technique requires measuring three or more interference patterns, each associated with a slightly different



**INTERFEROMETRIC IMAGE of a wear scar on a metal specimen shows the jagged surface terrain created when the specimen was stressed to its failure point. New, high-precision optical interferometric techniques developed by the authors are providing better understanding of how materials fail, thereby permitting the design of more reliable and longer-lasting products.**

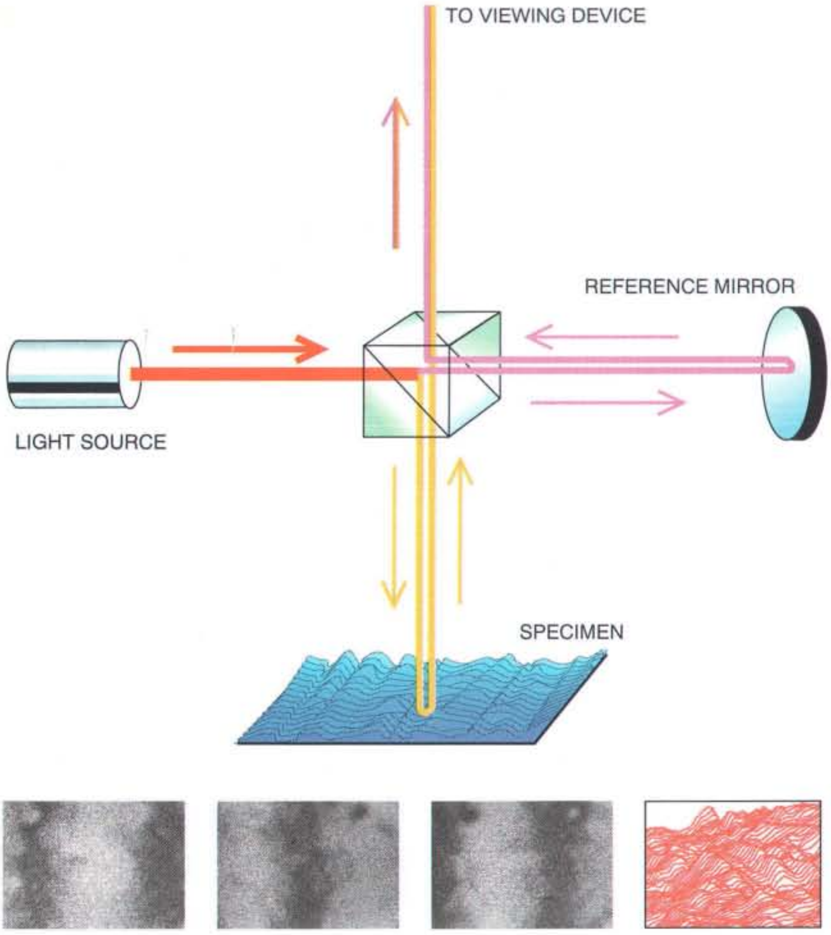
vertical position of the reference mirror or specimen. These positions must be within one wavelength (360 degrees of phase) of one another. Bruning realized that these three patterns together contain all the information necessary to determine the phase between the reference and specimen beams for every point on the sampled surface. Applying the appropriate trigonometric equations to the three patterns of light and dark yields a precise reading of the wave phases all across the sample surface. From this information, researchers can infer the surface topography that produced the observed phases.

## How Interferometry Works

Light interference occurs when two waves or sets of waves interact. In the prototypical Michelson interferometer (*below*), a laser beam strikes a half-silvered mirror, which splits the beam into two paths. One beam reflects off the specimen, the other off the reference mirror. When the beams recombine, waves that are out of step with one another partially or totally cancel one another out—hence the term "interference."

Repeating patterns of dark-to-light-to-dark, known as fringes, occur across the interference image, or interferogram (*bottom row*). Bumps and dips on the surface of the specimen change the length of the path traveled by the first beam, altering the spatial relation between the two beams and the shape of the fringes. Fringes deform around surface features on the specimen much like contour lines on a topographic map.

One approach to computer-analyzed interferometry involves making three interferograms of the specimen—in this case, a very rough videotape. Moving the specimen changes the distance traveled by the first beam and shifts the phases within the recombined beam. The varying brightness at each point of the interferogram is analyzed to reveal the wave phase, and therefore the surface height, at the corresponding point on the sample. That information is then plotted as a three-dimensional image (*bottom at far right*).



Measuring the relative phases of the reference and specimen beams gives far better vertical sensitivity than does the older method of measuring the displacement of the interference fringes in the reconstructed beam. The Bruning method also provides uniform coverage of the specimen's surface and makes it easy to automate the measurement process.

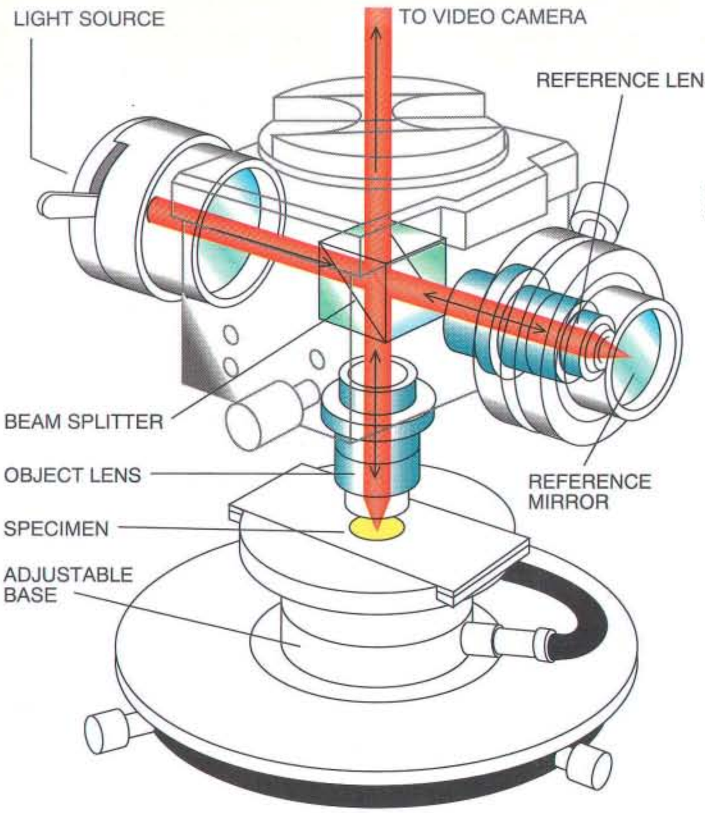
During the past decade, we have designed, built and refined two types of systems for performing direct phase-detecting interferometry. We did our work both at the 3M Company and at the department of physics at Bethel College and Seminary in St. Paul, Minn. One system, which embodies Bruning's technique, uses a Michelson or similar interferometer linked to an optical microscope to produce three-dimensional surface images. The other, less conventional system measures the two-dimensional surface profile of a specimen moving rapidly under a focused laser beam.

In the first kind of system the optical microscope magnifies the interference pattern produced by the interferometer. A video camera records this magnified image, converts it into digital form and stores it in a computer. The digitizing process divides the image into a grid. Each square of the grid defines one picture element, or pixel. Pixels are the smallest part of the interferometric image, somewhat analogous to the dots that make up a halftone newspaper photograph.

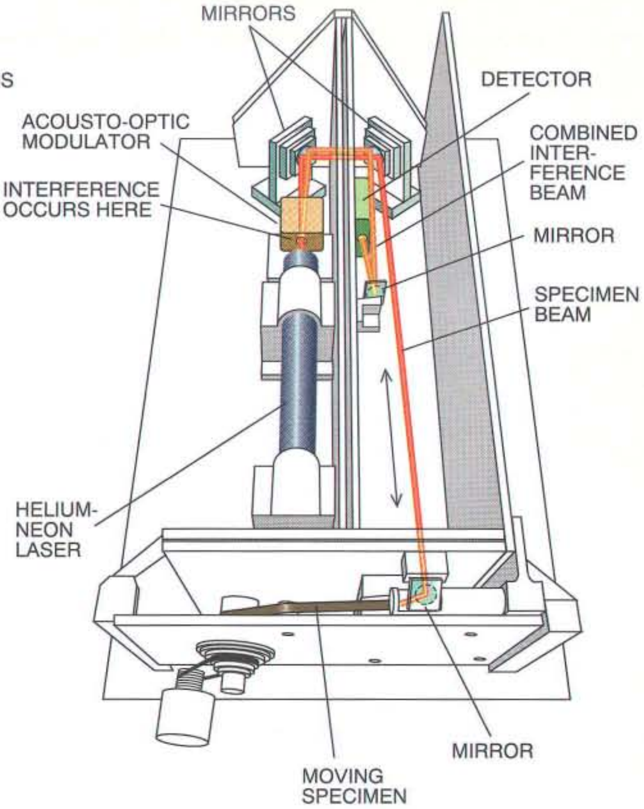
Moving the specimen or the reference mirror alters the phase at each pixel in the interference image. For the standard three-step Bruning method, the process is repeated twice in order to collect three interference images. Each time the specimen is moved one eighth of a wave, the round-trip path length changes by one quarter wave and the phase is shifted 90 degrees. The phase at each pixel of the interference image depends on the distance traveled by the light beam, which is affected both by surface texture and by the net movement of the specimen.

Laser-scanning interferometers detect wave phase as a function of time. Some of the first such devices were developed to measure rapidly changing thermonuclear plasmas. In our laser-based instruments, a beam is directed at a moving target such as a loop of magnetic tape or a rotating computer diskette. As with microscope-based devices, phase measurements are made by splitting and recombining the beam.

In a typical arrangement, laser-scanning interferometry begins with a beam



**MICROSCOPIC INTERFEROMETER is basically a Michelson interferometer that incorporates magnifying lenses and an adjustable specimen tray. A video camera collects the interference pattern for three sample positions and converts the patterns into digital data. Computer programs reconstruct these data into a three-dimensional image of the specimen's surface.**



**LASER-SCANNING INTERFEROMETER directs a focused laser beam on a moving specimen. A modulator shifts the frequency of the specimen beam, causing it to create a regular rhythm of interference when recombined with the emerging beam. Bumps and dips on the specimen change the path length, altering the rhythm and thereby revealing the surface topography.**

emitted by a helium-neon laser, which emits light having a 6,328-angstrom wavelength. In this case, it is more useful to consider the frequency of the light, that is, the number of wave cycles each second. A wavelength of 6,328 angstroms corresponds to a frequency of 474,100,000 megahertz (a megahertz is one million cycles per second).

The beam from the laser passes through an acousto-optic modulator, a device that alters the frequency of light. The emerging beam is shifted in frequency by 40 megahertz and is sharply focused onto a moving specimen. After reflecting from the specimen surface, the laser beam retraces its original path, passes once again through the modulator and is shifted an additional 40 megahertz. Finally, the beam reflects off the output laser mirror, where it recombines with the unshifted light emerging from the laser.

Two passes through the modulator alter the frequency of the laser light by 80 megahertz from its original frequency. When recombined with the original beam, the shifted light produces an 80-megahertz "beat" frequency of interference fringes. The two beams fall in and out of phase with each other 80 million times a second, and so the light intensity (brightness) at each point on the detector rises and falls 80 million times a second. Surface features on the specimen change the path length, and hence the travel time, of the reflected laser beam. That results in a constantly changing phase superposed on the regular rhythm of interference as the specimen moves under the laser beam. An electronic detector rapidly measures the phase and relays the information to a microcomputer, where it is analyzed to create a two-dimensional profile of the moving surface.

In one of the laser interferometers that we have designed, specimens such as videotape are formed into loops approximately 50 centimeters long and then slid over a polished guide located at the focus of the laser beam. Diskettes and other flat objects are moved under the laser beam on a turntable of another version of the laser interferometer. The laser scans an area one to 50 centimeters long and about one micron (millionth of a meter) wide. Specimens pass under the focused laser beam at 15 to 20 centimeters a second.

Both types of direct phase-detecting interferometers can resolve vertical surface features less than 10 angstroms high, approximately 30 times the precision possible using classical interferometric techniques. A microscopic interferometer designed to magnify the specimen 400 times yields a spatial resolution of about one micron and an imaged area measuring roughly 125 microns by 200 microns. Lowering the magnification reduces the resolution but increases the area of coverage.

For the laser-scanning interferometer, spatial resolution is a function of the distance between data points, which in turn depends on the speed at which the specimen passes under the beam and on the rate of data collection, or sampling. (Diffraction, a consequence of the wave nature of light, ultimately limits the resolution of both kinds of interferometry.) A typical distance between data points is 1.6 microns.

For most samples—videotape, for instance—the surface texture features of interest are about 10 microns in diameter, and the two interferometers produce comparable and complementary results. If the surface features are considerably longer, the laser-scanning interferometer gives a better measure of surface roughness because it samples an extended swath.

Collecting the requisite three interferograms for the microscopic interferometer takes approximately two seconds, about the same time needed to measure a surface profile with the laser-scanning interferometer. The computer requires about two minutes to analyze the surface based on the collected data.

The first step after processing the data is to compensate for phase errors. These occur when the topography of a surface rises or falls so sharply that the length of the light path changes by more than half a wave (180 degrees) between observations. Errors can also arise when a bump is so steep that light does not reflect back into the lens of the scanning device. In such cases, the computer loses track of the phase, because the phase repeats each cycle. For example, the crest of one wave looks exactly the same as the next one to the computer.

The relatively wide area imaged by the microscopic interferometer proves an advantage for phase error correction. Automated computer algorithms can detect mismeasured data and examine adjacent strips for comparison. In this way, the devices can go around the phase error without disconnecting parts of the image. Correcting for phase errors from the laser-scanning interferometer is more difficult because the beam examines only a narrow strip of surface, which does not provide any information about the surrounding area to serve as a reference.

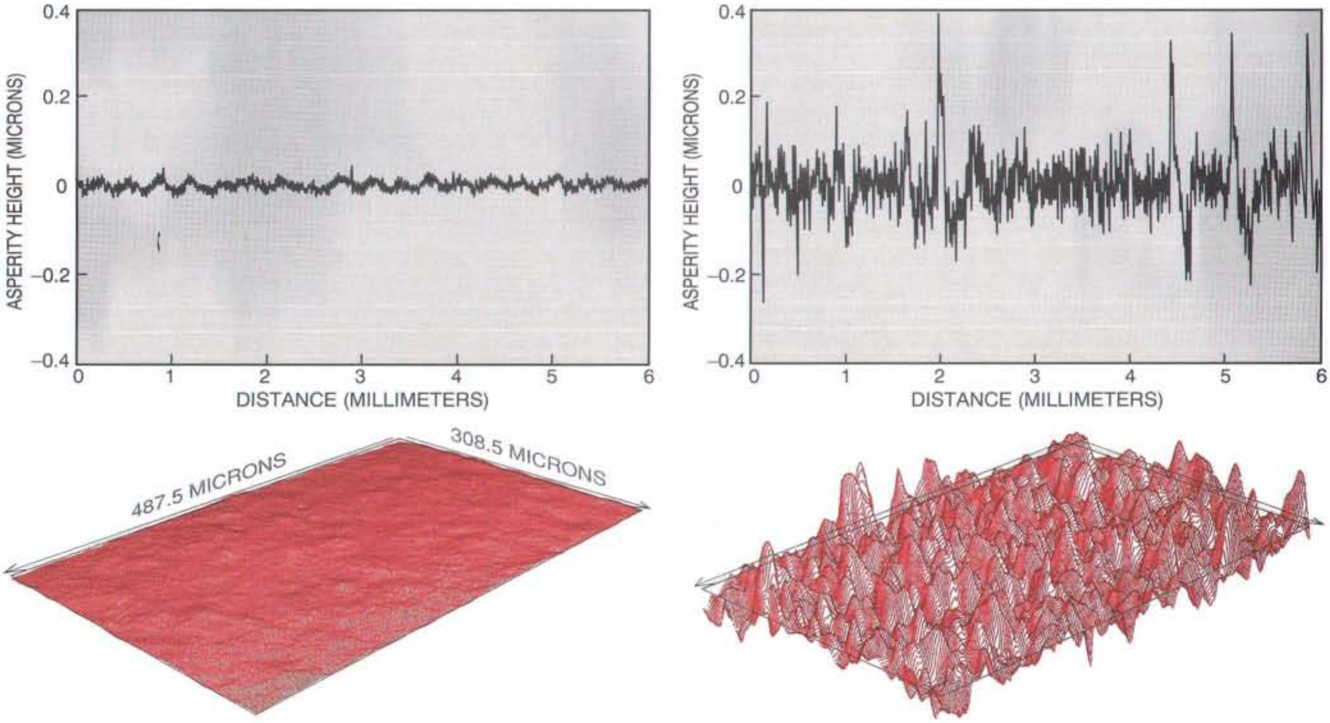
Data from microscopic interferometers must also be checked for curvature of the image, which can result from improper mounting, inherent specimen curvature or aberrations in the optical system. Statistical data-smoothing techniques are applied to find the curvature so that it can be mathematically subtracted to reveal only the surface texture that is of interest. Finally, corrected data are plotted in three-dimensional form.

Because the interferometric images are computer generated, they can be tilted, rotated and inverted. Graphics software can exaggerate the vertical relief of the image and add color coding to highlight topographic details. The software can also simulate an angled side view, a familiar and easily comprehended perspective. Optical and electron micrographs, in comparison, can be viewed only from an unnatural-looking vertical direction.

Computers can easily calculate the height, depth or volume of any surface feature detected by the interferometer. In addition, the overall surface topography within the area captured in the processed image can be analyzed by means of standard mathematical tools. For example, a plot of the number of surface irregularities versus their size, called a histogram, provides valuable information on how the surface texture of the sample will affect its performance. The average height of these irregularities and the standard deviation of their heights are also often useful.

Output from the laser-scanning interferometer undergoes similar processing. Once phase errors have been removed, the surface profile consists of slowly varying (low-frequency) and quickly varying (high-frequency) components. For relatively smooth specimens such as videotape, the high-frequency components contain most of the useful information. Low frequencies mostly represent instrumentation noise, vibrations and overall changes in sample thickness. Computer algorithms can effectively filter out these components. In the case of magnetic tape, however, low-frequency components may be of interest because they reflect variations in the thickness of the magnetic coatings.

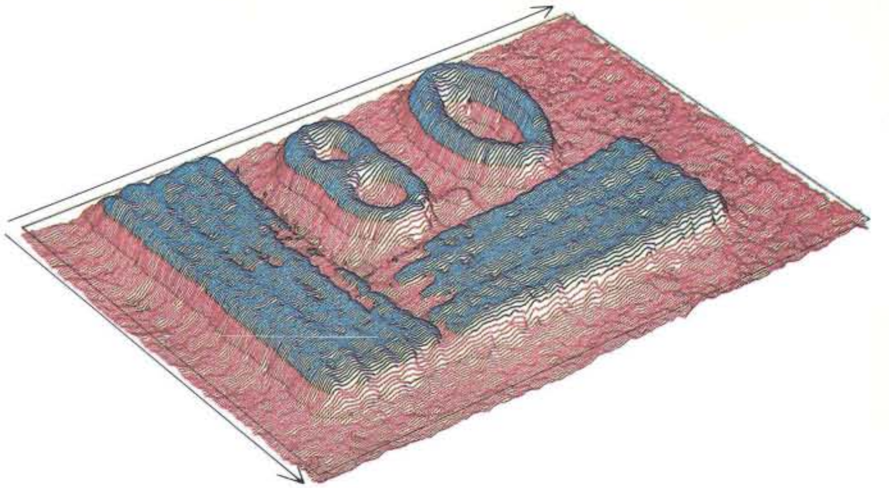
The linear surface profiles compiled by laser-scanning interferometers are more difficult to interpret than the highly graphic, aesthetically pleasing images produced by microscope-based de-



**MAGNETIC TAPE such as videotape must be extremely smooth to perform well. Interferometry makes it possible to measure the surface texture of the tape quickly and easily. Typical surface profiles of smooth and moderately rough magnetic tapes were measured using a laser-scanning interferometer (*top*) and a microscope-based device (*bottom*). Laser-scanning instruments generate long, two-dimensional surface profiles, which are useful for measuring the overall thickness of the magnetic coating on the tape; microscopic interferometers create more intuitively obvious three-dimensional images.**

vices. Statistical techniques condense the profiles and transform them into a more comprehensible and useful form. Such methods also enable researchers to measure the size and periodicity of repeated surface patterns. Statistical analysis can often reveal periodic surface features where a profile or three-dimensional interferometric image suggests only random order.

A number of researchers have developed alternatives to the interferometers that we have discussed. Frank E. Talke of the University of California at San Diego and David B. Bogy of the University of California at Berkeley have designed a laser interferometer that examines shifts in the frequency of laser light. Their goal is to study the motions of recording and playback heads caused by irregularities on the rapidly rotating surface of computer diskettes. James C. Wyant and Chris L. Koliopoulos of the University of Arizona have created interferometers similar to our microscopic interferometer.



**PHOTOGRAPHIC IMAGE of a test chart was examined using a microscopic interferometer. The topography of the film reveals the optical density of the image: the surface rises where the image is darker. By studying the surface texture of different film formulations, it is possible to analyze their potential for good resolution.**

Because of its unique capabilities, direct phase-detecting interferometry has moved rapidly out of the laboratory and into the commercial realm. Two- and three-dimensional direct phase-detecting microscopic interferometers are now commercially available from WYKO Corporation and Micromaps Corporation in Tucson, Ariz., and Zygo Corporation in Rochester, N.Y. Chapman Instruments, also of Rochester, sells a laser-scanning interferometer based on a different principle than the one described here. All four companies grew out of university research projects in interferometry.

At 3M, the direct phase technique is used routinely to characterize the surfaces of magnetic recording products, plastic films, recording heads and machined parts. Outside researchers also approach us periodically with various items they want to study. We have examined such disparate objects as photographic films, adhesive tape, wax on floor tiles, industrial calendering rolls (large cylinders used to press material into thin sheets), lenses and mirrors and even dental fillings.

Direct phase-detecting interferometry has played a particularly important role in refining the design and manufacture of magnetic tapes. Experiments show that the recording performance of a videotape correlates strongly with the roughness of the tape's surface. Laser-scanning interferometers can detect very slight periodic changes in the thickness of magnetic coatings. Because these instruments collect data quickly, in a noncontacting and nondestructive manner, we can monitor the actual recording performance of samples of freshly manufactured tape. With the help of interferometry, the cost of producing videocassettes, diskettes and other recording products has dropped significantly during the past decade. Performance and quality of these items also have sharply improved.

Learning how products like computer diskettes or ball bearings wear out is necessary for developing better versions. Unfortunately, the amount of material lost to wear is usually minute and localized, even when these products are driven to failure. Interferometric techniques have improved measurements of rates of wear, calculations of the volumes of wear tracks and identification of the mechanisms of wear.

In processed photographic films, surface topography often follows the optical density (that is, darkness) of the image. Interferometry helps to establish the sharpness of photographic images by measuring the slope of the surface topography of the image. Such information is valuable for the evaluation of new film formulations.

The advent of direct phase-detecting interferometry has made the measurement of the surface texture of polyester films routine. Polyester is used as a backing in a broad range of applications, including adhesive tape, simulated wood-grain trim, magnetic tape and photographic film. Often polyester film is given a textured surface on one side so that it will unwind easily from huge stock rolls. The other side, however, must be extremely smooth, or else the film will not work as a backing. Previously it was difficult to determine quantitatively the degree of texture applied to the material. Now interferometry is permitting better control of the texture of polyester films and hence is leading to the development of superior polyester-based products.

Primarily as a result of tremendous leaps in computer power and the capabilities of graphics software, optical interferometry can now be used to perform high-resolution, two- and three-dimensional surface mapping. Adapting these techniques to surface studies has just begun. Nevertheless, we are convinced that they are destined to become important tools for both industry and academia.

FURTHER READING

Digital Wavefront Measuring Interferometer for Testing Optical Surfaces and Lenses. J. H. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White and D. J. Brangaccio in *Applied Optics,* Vol. 13, No. 11, pages 2693–2703; November 1974.

Light and Its Uses: Readings from *Scientific American.* With introductions by Jearl Walker. W. H. Freeman and Company, 1980.

Interferometric Measurements of the Surface Profile of Moving Samples. R. W. Peterson, G. M. Robinson, R. A. Carlsen, C. D. Englund, P. J. Moran and W. M. Wirth in *Applied Optics,* Vol. 23, No. 10, pages 1464–1466; May 15, 1984.

Three-Dimensional Surface Metrology of Magnetic Recording Materials through Direct-Phase-Detecting Microscopic Interferometry. D. M. Perry, P. J. Moran and G. M. Robinson in *Journal of the Institution of Electronic and Radio Engineers,* Vol. 55, No. 4, pages 145–150; April 1985.

# Biological Control of Weeds

*Insects and microorganisms are already serving as commercial weed killers, and other biological approaches show promise. The goal: environmentally compatible alternatives to chemical herbicides*

by Gary A. Strobel

Humans have struggled against weeds since at least the beginning of agriculture. Some people would even say the fight is literally as old as Adam. In Genesis 3:18, one of the earliest references to the noxious plants, Adam is promised thorns and thistles in his otherwise perfect Eden.

Marring gardens is among the milder effects of weeds—any plants that thrive where they are not wanted. Weeds also clog waterways, destroy wildlife habitats and impede farming. Their rampant spread eliminates grazing areas and accounts for a third of all crop losses in the world, devastation the expanding human population can ill afford. Crops and other desirable plants may die because the interlopers compete for water, nutrients and sunlight or interfere with irrigation. Harvesting may be hampered as well.

The global need for weed control has been answered mainly by the chemical industry. Its herbicides are often effective and necessary to agriculture. Yet some pose serious problems, particularly if they are misused. For example, toxic and otherwise harmful compounds threaten animal and public health when they accumulate in food plants, groundwater and drinking water. They can also directly harm the workers who apply them.

In recent years the chemical industry has introduced several herbicides that are more ecologically sound than those of the past, notably glyphosate, produced by Monsanto, and a series of compounds called sulfonylureas developed by Du Pont. Yet new chemicals alone will not solve the world's weed problems. Hence, an increasing number of scientists, including my colleagues and me at Montana State University, are exploring biological alternatives that harness the innate weed-killing powers of living organisms, primarily insects and microorganisms.

GARY A. STROBEL is Richard Gray Professor of Plant Pathology at Montana State University. He is also director of the state's Experimental Program to Stimulate Competitive Research, a project sponsored by the National Science Foundation to improve funding in states that receive relatively little federal research support. Some readers may recall that in 1987, without having secured federal approval, Strobel inoculated 14 elms with a genetically modified bacterium designed to protect the trees from Dutch elm disease. The move was in part a protest against what he regards as excessive bureaucracy. He believes the incident helped to prompt simplification of the Environmental Protection Agency's guidelines for biotechnology research. This is his third article for SCIENTIFIC AMERICAN.

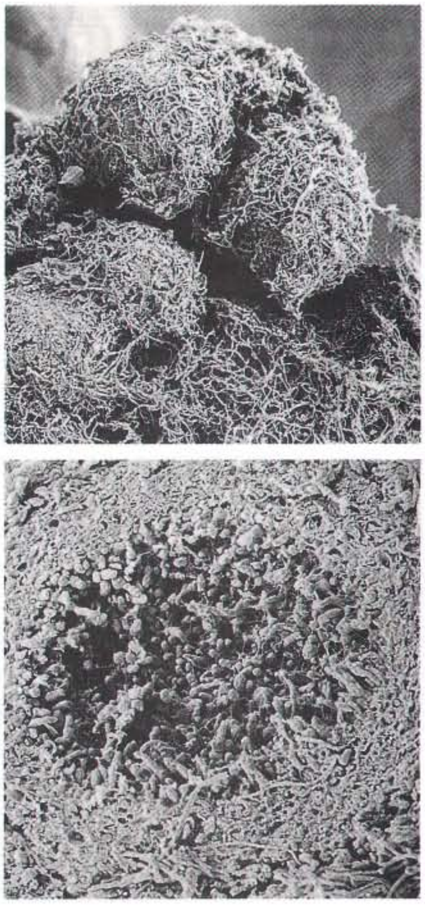
The biological agents now in use are environmentally benign, and many offer the added benefit of specificity. They can usually be chosen for their ability to attack selected targets, leaving crops and other plants untouched (including plants that might be related to the target weeds). In contrast, some of the most effective chemical herbicides kill virtually all plants they contact, sparing only the few species possessing natural resistance to the chemicals and species that have been genetically modified for resistance. Furthermore, a number of biological agents can be administered once, after which no added applications are needed. Chemicals typically have to be applied several times in a growing season.

Biological approaches might never supplant chemical herbicides altogether, but they should help limit the use of chemicals and thereby reduce the associated risks. They might also make it possible to conquer weeds that defy management by standard herbicides.

Of course, people are the original biological weed-control agents. For thousands of years, they have pulled, cut, tugged, slashed and stomped their way through weed-infested land, with uneven effect.

The first successful introduction of a nonhuman biological agent is believed to have occurred in the mid-1800s. By then, the cactus *Opuntia vulgaris*, a form of pricklypear native to the Americas, had made a nuisance of itself in India. It was finally controlled by release of the insect *Dactylopius ceylonicus*, a natural enemy of the plant.

Since that time, insects have been deployed to limit proliferation of more than 30 weed species in various parts of the world. This work has paved the way for modern experimentation with



**DISCOLORED LEAF on purple nutsedge (*Cyperus rotundus*) is the handiwork of a fungus, *Ascochyta cypericola*, one of**

microorganisms that kill or debilitate weeds. Several such pathogens are already on the market or poised to enter it, and dozens of others are under intense scrutiny in many laboratories. Most of the microbes being studied are fungi, because these are the most common plant pathogens, but a few groups of workers around the world are also examining plant viruses, bacteria and nematodes (roundworms).

To develop biocontrol measures, investigators must learn a great deal about each target weed. In particular, they need to know its natural enemies and its life cycle, so that they can deliver the most potent enemies at the time in the cycle when the agents are particularly destructive.

The chemical industry is now trying to develop herbicides in an equally rational manner, designing certain compounds to interfere with specific chemical processes in selected weeds. In the past, however, many standard herbicides were born of a kind of tinkering. That is, chemists created novel compounds with no particular biochemical target in mind; they then screened the compounds against a range of plants.

Typically the search for enemies begins with a review of the scientific literature on the weed, but it may also involve scouting fields for previously unidentified antagonists. The hunt often begins at the original home of the weed, because its enemies tend to be most prevalent there. Many plants that become odious in one place are tame at their center of origin, precisely because other organisms that have evolved in the same region suppress their spread.

For example, expeditions to sites of origin have turned up promising biological agents against purple nutsedge (*Cyperus rotundus*), a grasslike plant that sprouts lavender flowers. After escaping from India, it went on to infiltrate crops in tropical regions throughout the globe. A persistent pest with no redeeming features, it is widely considered to be the world's worst weed.

Other examples abound. Workers venturing abroad have also discovered natural enemies of such weeds as wild oats (*Avena fatua*), which spread from the Middle East to virtually all grain-growing areas on the earth; Johnsongrass (*Sorghum halepense*), a Mediterranean export that is a bane of sorghum and corn crops in most agricultural areas; marijuana (*Cannabis sativa*), a native of temperate Asia; and such widespread aquatic nightmares as water hyacinth (*Eichhornia crassipes*), which originated in South America [see "Waterweed Invasions," by Spencer C. H. Barrett; SCIENTIFIC AMERICAN, October 1989]. In every case, accidental or misguided export of these plants has cost millions of dollars in crop losses or control measures, or both.

Whether at the initial or adopted home of a plant, researchers identify potential biocontrol agents by looking for target plants that display symptoms of disease or infestation by insects. Common signs include yellowed or dying leaves, rotting flower parts, decayed roots or damaged stems (in the case of infection) or the presence of eggs or actual insects (in the case of infestation). Usually the agents of destruction can then be isolated or else derived from spores or eggs they leave on the plants.

Finding such agents is surprisingly easy. For instance, just last year my colleague Rajeev K. Upadhyay and I dis-

**many potential biological herbicides. The fungus proliferates on the weed by generating pycnidia (*rounded bodies in top scanning electron micrograph*), each of which contains spores (*swollen tips on threads in cutaway*). The pycnidia are magnified 31 times; the spores, 305 times. Wilford Hess of Brigham Young University made the micrographs.**

# High Speed Telecommu Philips makes O

All over the world, Philips sells and implements telecommunication systems and networks for public operators and private companies.

Providing innovative solutions in key areas such as radio and cable transmission, network access, mobile communications, rural telephony, data and business communications.

A prestigious example is the digitalisation of the

# nications for Spain in '92. lympic History.

communications facilities for the '92 Olympics in Barcelona. Here Philips is making Olympic history by installing, for the Spanish PTT, the world's first high speed, high volume digital network using SDH, the new world transmission standard.

Philips is also installing a digital telephony system at Barcelona and Malaga airports and is the official supplier to the Olympics of mobile radio terminals.

These winning state-of-the-art solutions are striking proof of Philips' commitment and dedication to the market.

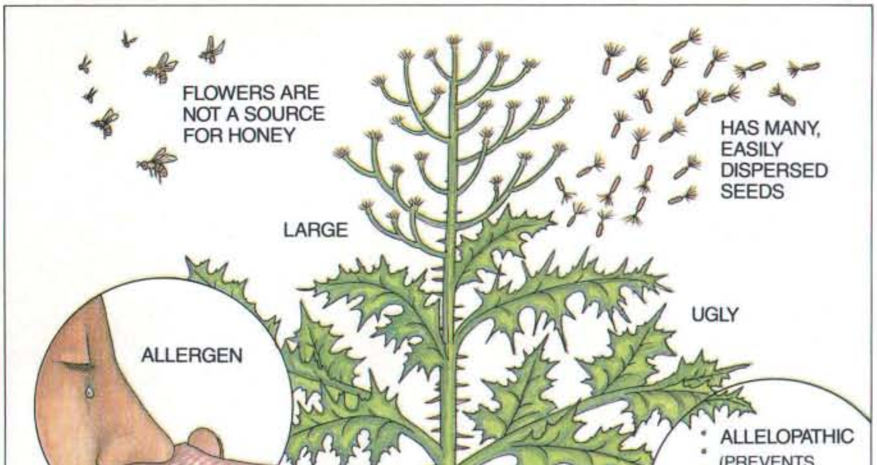
Philips. In perfect shape today, to further shape the future of telecommunications tomorrow.

Excellent reasons to go to Geneva in October to see Philips at Telecom '91. For further information call free on (UK) 0800 585509.

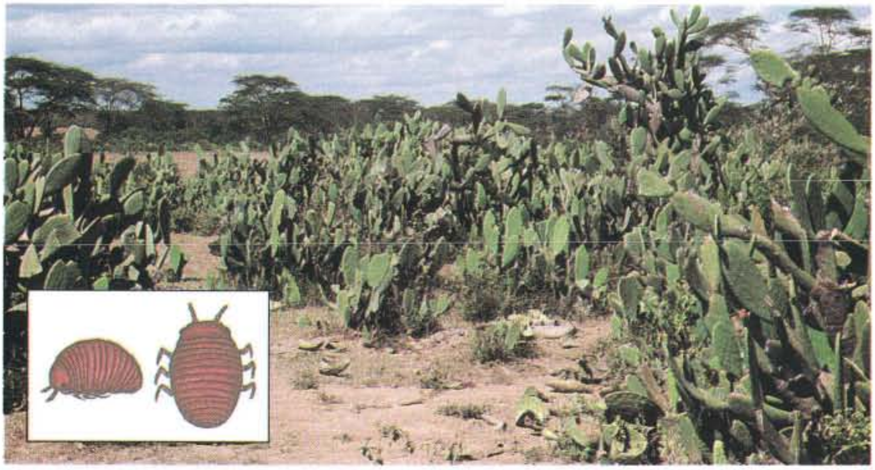
Telecommunications, talk with Philips!

PHILIPS

**PHILIPS**


FLOWERS ARE NOT A SOURCE FOR HONEY
HAS MANY, EASILY DISPERSED SEEDS
LARGE
UGLY
ALLERGEN
ALLELOPATHIC (PREVENTS OTHER PLANTS FROM GROWING)
FORMS ROSETTES
SPINY
UNPALATABLE; NO FOOD VALUE FOR WILDLIFE
SPREADS QUICKLY
SEEDLINGS GROW RAPIDLY
RESISTS REMOVAL
TAPROOT
RESISTS HERBICIDES
TOXIC

**ULTIMATE WEED has no redeeming features. Unsightly and useless, it is unpalatable or toxic to animals and devoid of sugar for honey. It grows tall, proliferates rapidly and actively inhibits the development of other plants (allelopathy). It also triggers allergic reactions in humans. And it won't go away: with its long taproot and rosettes, it resists mechanical removal, and it defies elimination by herbicides.**



**CACTUS *Opuntia vulgaris* was the first weed to be controlled by an insect, in this case, *Dactylopius ceylonicus* (*inset*). The time was the mid-1800s; the place, India.**

covered a previously unrecognized fungal pathogen of purple nutsedge simply by searching in India for plants displaying yellowed or necrotic spots. We are now examining the possibility that the fungus, *Ascochyta cypericola*, can be exploited to manage that weed.

In general, biologists focus further study most closely on agents that specifically and efficiently attack a single weed or a limited range of related weeds. The organisms will not be released freely, however, until several steps have been followed carefully. These include greenhouse tests, followed by limited field trials in small, isolated plots and, subsequently, more extensive field studies.

In the case of foreign insects and pathogens, federal and state laws in the U.S. and in many other countries dictate that investigators carry out all laboratory analyses and greenhouse tests under quarantine conditions, restricting the entry and exit of people, air and water. Indeed, workers must handle foreign organisms with extra care at every stage of study to avoid causing unwanted epidemics. In the past, tourists and other travelers have caused such epidemics by accidentally unleashing the microbes responsible for Dutch elm disease, chestnut blight and diseases of citrus fruits, apples, pears and bananas.

The organisms that industry or government agencies ultimately adopt for weed management typically have to meet several criteria. Insects must be able to reproduce readily. Likewise, pathogenic microorganisms have to be able to generate abundant spores or other structures that give rise to new generations. Microbes must also remain viable when they are "formulated": mixed with substances that prevent desiccation and facilitate adhesion to weeds. And, except when one wants to attack a broad spectrum of weeds, the organisms should be specific to a target weed and able to debilitate it under a range of environmental conditions. In the U.S., regulations also require investigators to produce evidence of safety before an organism can be tested outside the laboratory and distributed on a wide scale.

The first pathogenic organisms to pass such scientific and regulatory hurdles began to join the more established insect-based products on the market in the early 1980s [*see table on page 56D*]. Some of the products now available or likely to be sold in the near future are indigenous to the area of application; others are imported from the weed's original home.

# IF YOU THINK WE MAKE CARS ONLY, YOU'RE 21% CORRECT.

You probably think of Hyundai as the world-class maker of affordable high quality automobiles. And you're right of course. But Hyundai Motor represents only about 21% of the Hyundai Business Group. Shipbuilding, electronics, general construction, machinery and heavy equipment, rolling stock and petrochemicals are all key components of the Hyundai group.

Hyundai started as a small civil contractor 44 years ago, and today we've over US$20 billion worth of overseas engineering and construction experience. Hyundai also built and operates a dockyard with the largest ship tonnage output in the world and has won awards for building the World's Best Ship for eight consecutive years. Those who are involved with computers, semiconductors, and telecommunications have also come to know and trust Hyundai.

In fact, corporate depth and diversity like this is one reason why Hyundai has become such a symbol of Korea's economic success. If you think Hyundai is the right partner for you, join with us in Korea or around the world as we bring the challenge of tomorrow closer to today.

HYUNDAI

K.P.O. Box 92 Seoul. Korea Tel: (02)746-1873
Tlx: K23175/7 Fax: (02)741-2341

The performance of one of the pioneers in the commercial world, a strain of the fungus *Colletotrichum gloeosporioides* (trade name: Collego), is in certain ways typical of the new pathogen-based herbicides. This particular fungus, a representative of the indigenous group, comes from the American South and is helping limit the damage done by northern jointvetch (*Aeschynomene virginica*) in rice fields of Arkansas. It was identified as a potential control agent and formulated and tested by George E. Templeton and his group at the University of Arkansas.

The chemical compounds—phenoxy herbicides—that farmers normally deploy against the weed kill essentially every northern jointvetch plant, but when the chemicals are applied carelessly, they can damage rice as well as cotton and soybeans in nearby fields. In contrast, the "mycoherbicide" does not harm crops. It does take longer than standard chemicals to show an effect, and it is slightly less potent, killing perhaps 99 percent of its targets. But the time span is reasonable, and the kill rate is acceptable, because eradication of every last weed is not necessary if the population is reduced enough to prevent the weed from competing with desirable plants.

The rust fungus *Puccinia chondrillina,* which actually causes plants to look rusty, has a similar track record. An example of an imported pathogen, it is helping Australians to restrain skeletonweed (*Chondrilla juncea*), a Mediterranean export. As before, this mycoherbicide does not eradicate every skeletonweed plant, but even so, it is saving millions of dollars a year—money that would have been spent on traditional herbicides and labor.

Unfortunately, as is true in many areas of science, early promise does not always translate into useful applications. Indeed, most insects and microorganisms that keep weeds in check in the greenhouse turn out to be ineffective in the field, where both climate and the properties of soil are more variable and unanticipated natural enemies of the biological agent may roam. For example, Peter Harris of Agriculture Canada (the Canadian equivalent of the U.S. Department of Agriculture) has noted that in Australian field tests, only five insects out of 51 introduced to con-

## A Sampling of Weed Pathogens on the Market or under Study

| PATHOGEN | WEED | WHERE WEED IS TROUBLESOME |
|---|---|---|
| **Now Sold Commercially** | | |
| *Colletotrichum gloeosporioides*<br>trade name: Collego | Northern jointvetch<br>(*Aeschynomene virginica*) | Rice crops in Arkansas |
| *Phytophthora palmivora*<br>trade name: DeVine | Milkweed vine<br>*(Morrenia odorata)* | Citrus orchards in southeastern U.S. |
| **Expected to be Marketed** | | |
| *Alternaria cassiae*<br>trade name: Casst | Sicklepod<br>(*Cassia obtusifolia*) | Soybean and peanut crops in southeastern U.S.; various crops in Australia |
| *Colletotrichum gloeosporioides*<br>(different strain from that used in Collego)<br>trade name: BioMal | Roundleaf mallow<br>(*Malva pusilla*) | Small-grain crops in U.S. and Canada |
| **Showing Promise in Field Tests *** | | |
| *Alternaria helianthi* | Cocklebur<br>(*Xanthium pennsylvanicum*) | Soybean crops in southern U.S.; crops in Australia and Israel |
| *Alternaria macrospora* | Spurred anoda<br>(*Anoda cristata*) | Cotton and soybean crops in southern U.S. |
| *Cercospora rodmanii* | Water hyacinth<br>(*Eichhornia crassipes*) | Waterways in tropical and semitropical parts of the world |
| *Colletotrichum coccodes* | Velvetleaf<br>(*Abutilon theophrasti*) | Corn and soybean crops in midwestern U.S. |
| *Sclerotinia sclerotiorum*<br>(genetically modified) | Canada thistle<br>*(Cirsium arvense)* | Yards, pastures, ditch banks and small-grain crops in temperate parts of the world |

* Selected from a large list of promising organisms.

Northern jointvetch

Milkweed vine

Sicklepod

Roundleaf mallow

Cocklebur

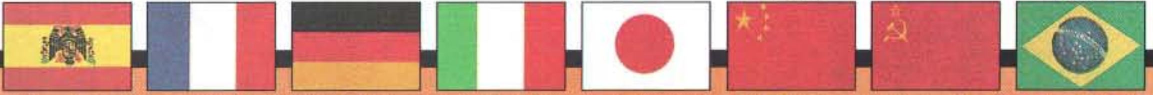
Spurred anoda

Water hyacinth

Velvetleaf

Canada thistle

| SPANISH | FRENCH | GERMAN | ITALIAN | JAPANESE | Mandarin CHINESE | RUSSIAN | Brazilian PORTUGUESE |
|---|---|---|---|---|---|---|---|
| 30 Cassettes | 30 Cassettes | 30 Cassettes | 30 Cassettes | 30 Cassettes | 30 Cassettes | 30 Cassettes | 30 Cassettes |
| + Triple Bonus | + Triple Bonus | + Triple Bonus | + Triple Bonus | + Triple Bonus | + Triple Bonus | + Triple Bonus | + Triple Bonus |
| $245.00 | $245.00 | $245.00 | $245.00 | $265.00 | $265.00 | $265.00 | $265.00 |



# Learn Foreign Languages... Incredibly Fast!

**Conversing in a foreign language is a major social and business asset...and brings new life to the worlds of travel, entertainment, and relationships. The technique of *accelerated learning*, as conveyed by these proven foreign language courses, allows anyone to comfortably converse in a new language within 30 days.**

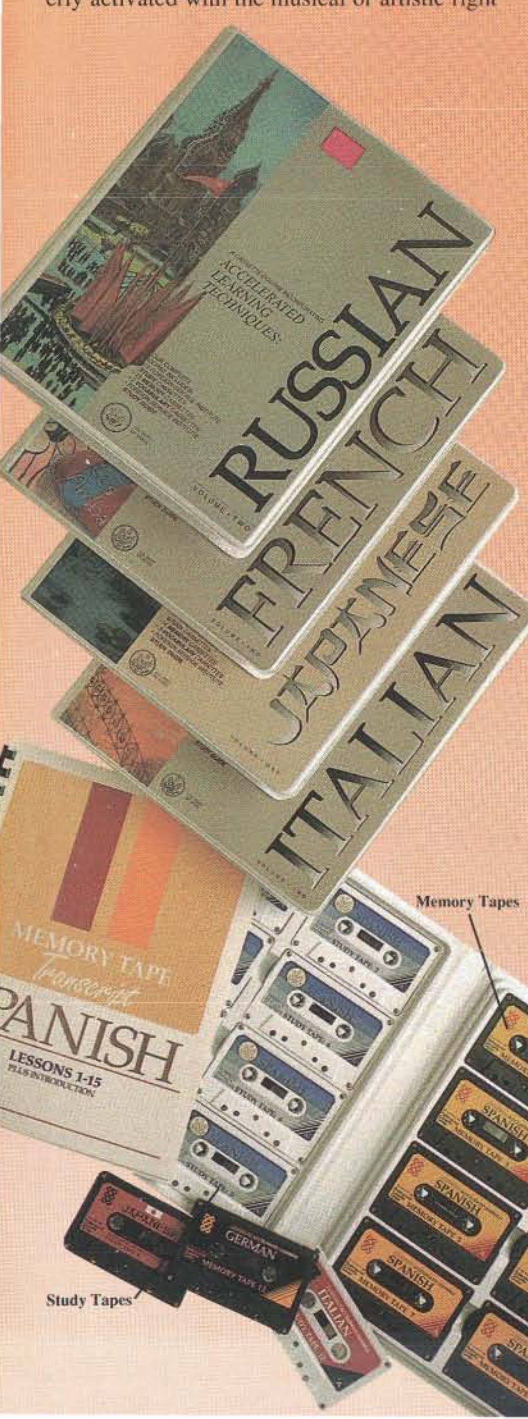
Accelerated learning, developed by famed learning expert **Dr. Georgi Lozanov,** is based on the premise of involving both hemispheres of the brain in the education process. The analytical or logical left side of the brain, when properly activated with the musical or artistic right side of the brain, both increases the speed and heightens the retention of learning. Utilizing these untapped mental capacities of your learning ability is the basis of this unique, highly effective course.

You *will learn* the language as stresslessly as a child does, by hearing new vocabulary and phrases in alternately loud, whispered, and emphatic intonations, all accompanied by slow rhythmic music in digital stereo. This perfect combination of music and words allow the two halves of the brain to work together to dramatically facilitate your assimilation of the new language.

The first 15 (memory) tapes of this 30-tape package help activate the learning capacities of the brain. The second 15 (study) tapes are the very same tried and proven tapes used by the **Foreign Service Institute** to train career diplomats. This marriage of two concepts literally gives you two courses in one, providing the best of both worlds in language instruction.

Best Value! With a total of 32 cassettes plus study materials, this program represents the best value available today in language instruction. Compared to other programs, the **Accelerated Learning Series** outperforms them with twice the audio and 20 times the study material.

To correctly converse in a foreign language, you must understand the meanings and intent of the native speaker. If, after 30 days of listening to the study and memory tapes, you are not *comfortably understanding and conversing in your new language, return them for a full refund.*



**"American Managers with Language Skills Open More Doors"**

*–Wall Street Journal Editorial*

**"Company and marketing executives will find after 1992 that it is a handicap not to be fairly conversant with at least one other major European language – and preferably two or three..."**

*–The London Times*

Triple Bonus !!

You'll also receive:

- Two 90-minute Vocabulary Tapes
- The 100-page *How To Learn A Foreign Language*
- The *American Express International Traveler's Dictionary*

TO ORDER: Phone or send your check, money order or Inst. P.O.

TOLL-FREE 24 HRS: VISA · M/C · AM. EX.

**1•800•85•AUDIO**

Rush Orders PHONE 9-5 PDT:

**1•818•799•9000**

You may FAX your credit card order or company P.O. to:

**1•818•792•7815**

**• INTERNATIONAL ORDERING INFORMATION •**

"**New!** Now, for your ordering convenience, you may call our order desk toll-free 24 hours a day from any of the following countries via **AT&T International 800 Service.**"

| | | | |
|---|---|---|---|
| BELGIUM | 11-6599 | NETHERLAND | 06-022-4612 |
| DENMARK | 8001-0578 | SINGAPORE | 800-1625 |
| FRANCE | 05-90-1368 | SPAIN | 900-98-1120 |
| GERMANY | 0130-81-1139 | SWEDEN | 020-793-626 |
| ITALY | 1678-70-179 | SWITZ | 046-05-9632 |
| JAPAN | 0031-11-1907 | UK | 0800-89-7452 |

☐ **FRENCH $245.00**
☐ **SPANISH $245.00**
☐ **GERMAN $245.00**
☐ **ITALIAN $245.00**
☐ **JAPANESE $265.00**
☐ **CHINESE (Mandarin) $265.00**
☐ **RUSSIAN $265.00**
☐ **PORTUGUESE (Brazilian) $265.00**

Name

Address

City State Zip

Credit Card No Exp.

Signature (Card Orders Only)

☐ VISA ☐ MASTERCARD ☐ AMERICAN EXPRESS

Need It Tomorrow? Ask Operator for **Express Service**

FEDERAL EXPRESS

**PROFESSIONAL CASSETTE CENTER**
350 WEST COLORADO BOULEVARD
DEPARTMENT SCI
PASADENA, CALIFORNIA 91105 U.S.A.

Please add $11.00 shipping & handling
California residents add 7% sales tax.

**DEAD PATCH (*left*) in a field of spurred anoda (*Anoda cristata*), a weed common in the eastern half of the U.S., demonstrates the herbicidal potential of the fungus *Alternaria macrospora*. Such limited field tests are required before larger tests of microbes or insects can be attempted. Specimen of spotted knapweed (*Centaurea maculosa*) (*left plant in photograph at right*) fell victim to a different biological strategy. It was damaged by maculosin, a knapweed-specific phytotoxin (plant-killing substance) extracted from the fungus *Alternaria alternata*. An untreated specimen remained healthy.**

trol pricklypear have proved helpful.

On the other hand, the accomplishments to date demonstrate that biological agents can in fact be effective and serve as alternatives to chemicals. Furthermore, if the protocols I have described are followed, the risks of mistakenly harming desirable plants are small. Suzanne W. T. Batra of the Agricultural Research Service at the U.S. Department of Agriculture has pointed out that of more than 100 insects introduced for weed management worldwide, none has altered its eating habits and abandoned its host weed in favor of an economically important plant. Neither has there been any scientific misjudgment of the pathogens that have been widely released; not one has suddenly begun attacking plants it had not harmed before.

My co-workers and I are studying a biological approach that would bypass the need to release whole organisms and would thus erase even the minuscule risk that pathogens might unexpectedly take up residence in new hosts. In place of organisms, the strategy relies on substances the microbes produce, namely, weed-damaging compounds known as phytotoxins. Application of such compounds would also obviate the need to ensure that microbes remain viable after they are formulated and put in storage or applied to weeds.

The study of phytotoxins has appeal for other reasons as well. After investigators extract them from pathogens, the toxins can be studied individually for clues to exactly how the pathogens kill weeds. As we decipher the structures of the compounds, we should also be able to synthesize these molecules, thus escaping the need to collect or maintain colonies of pathogens to produce weed-killing formulations. What is more, we should be able to synthesize many derivatives to improve the effectiveness of the original toxins.

My laboratory has been studying weed-directed phytotoxins for 10 years in collaboration with Jon C. Clardy of Cornell University, Fumio Sugawara of the Institute of Physical and Chemical Research near Tokyo and a host of students and postdoctoral fellows at both locations. A decade is actually a relatively short time, considering that the work is extremely time-consuming and that only our group and a handful of other laboratories are working on phytotoxins. Nevertheless, important progress has been made.

More than 25 weed-killing fungi have been examined. The studies show that practically all such fungi, except perhaps rusts and mildews, make phytotoxins. When they produce such a toxin, they also apparently produce a family of compounds that are structurally related to it. In any family the mixture may include both nontoxic and phytotoxic molecules.

At times, related compounds kill by acting in concert, a finding that suggests some phytotoxin-based herbicides of the future will consist of several members of a single structural family. What, though, is the role of substances that have no toxicity? I suspect fungi sometimes generate many derivatives of a single molecule to assure that at least one of them will harm the plant and thus guarantee that the disease-causing fungus will have access to the plant's nutrients.

In a mildly disconcerting result, only one phytotoxin so far isolated from any weed-specific fungus is itself specific to the host weed (an indication that the specificity of most fungi derives from other factors). Some of the phytotoxins are at least selective: they harm a limited set of related weeds. Yet most others have no selectivity and could potentially damage valued plants if they were applied as herbicides. Still, discovery of even a few specific or selective phytotoxins implies that others exist; they merely await discovery. Indeed, if one assumes that there are thousands of weed species in existence and that each is sensitive to perhaps 20 or more pathogens, then the list of fungi and other pathogens still to be examined is quite enormous.

During the past decade, the structures of a number of phytotoxins and their relatives have been elucidated. Because research into the weed-killing powers of phytotoxins is relatively new, no one yet knows how most of the compounds damage plants. In one exception, however, my colleague Doug Kenfield, here at Montana State, has explained why triticones A and B, which derive from the fungus *Curvularia clavata*, are toxic to several grasses. They inactivate enzymes critical to the survival of the weeds by combining with sulfhydryl groups (sulfur and hydrogen) on the active sites of the enzymes. So bound, the enzymes cannot interact effectively with other molecules.

My group's basic approach to identifying and analyzing phytotoxins, and the challenges and excitement associated with that work, can best be described by the example of maculosin, a phytotoxin that affects spotted knapweed (*Centaurea maculosa*). This weed, which produces pinkish flowers dotted with white spots, plagues the northwestern U.S. and southwestern Canada.

The plant was probably introduced in the 1920s, after its seeds became mixed with others imported from Europe, such as those of alfalfa. Beekeepers traveling from Europe may have introduced the plant as well, intending it to serve as a source of honey. The insects do make honey from the nectar in the flowers, but the presence of sugary material is perhaps the plant's only redeeming feature.

Knapweed readily displaces most grasses and flowering plants in its new home, quickly taking over hillsides and pastures. The lands become worthless for grazing because animals will not eat the weed. Indeed, its infiltration of rangeland costs ranchers million of dollars each year in Montana alone—an amount that is growing along with knapweed's dominion.

In the spring of 1984 a student named Andrea C. Stierle told me she wanted to combine the study of chemistry and plant pathology in her doctoral work. I urged her to try to isolate and analyze phytotoxins active against knapweed. The task would be difficult: a scan of the weed literature revealed that not one pathogen had yet been described for the plant.

Stierle and her husband spent the summer vainly searching the mountain pastures of Montana for diseased specimens. Then, one day early in the following September, on a hillside near her home in Butte, she unexpectedly noticed a seriously diseased knapweed plant, its leaves and stems marred by black spots indicative of tissue death. Back in the laboratory, she isolated a fungal pathogen from the lesions and showed that it caused similar spots on healthy knapweed plants in the greenhouse. The pathogen turned out to be a new strain of a common fungus, *Alternaria alternata.*

Now Stierle turned her attention to chemistry. In the laboratory of John H. Cardellina II at the university, she grew cultures of the fungus and laboriously extracted and purified the compounds it produced. Then she determined the chemical structure of the molecules that seemed damaging to knapweed.

Some of the active products were phytotoxins known to be made by other fungi as well. Yet two related products were unfamiliar. One of these was only mildly toxic, but the second, which we called maculosin, amazed us. Even

## A Selection of Phytotoxins under Study for Weed Control

The author's group recently isolated the compounds below (from fungi infecting the "host" weeds listed) and deciphered their structures. The team is analyzing the herbicidal potential of the listed phytotoxins and many of their close relatives as well as of a number of other phytotoxins. Molecules that are host specific (efficiently killing a single weed species) or that selectively damage a limited range of species hold the most promise for commercial development. Regrettably, as the sampling here shows, such choosy compounds seem to be relatively rare.

| PHYTOTOXIN | SOURCE FUNGUS | HOST WEED | WHERE WEED IS TROUBLESOME | COMMENT |
|---|---|---|---|---|
| Bipolaroxin | *Bipolaris cynodontis* | Bermudagrass (*Cynodon dactylon*) | Corn, cotton, sugarcane and other crops worldwide | First-known host-selective phytotoxin |
| Curvulin | *Drechslera indica* | Common purslane (*Portulaca oleracea*) | Dozens of crops worldwide | Host selective and works well at low concentrations |
| Cypreine | *Ascochyta cypericola* | Purple nutsedge (*Cyperus rotundus*) | Dozens of crops worldwide | Not selective but extremely toxic to purple nutsedge |
| Exserohilone | *Exserohilum holmi* | Crowfoot grass (*Dactyloctenium aegyptium*) | Corn, cotton, sugarcane and other crops in tropics and semitropics | Not selective |
| Gigantenone | *Drechslera gigantea* | Bermudagrass (*Cynodon dactylon*) | Corn, cotton,sugarcane and other crops worldwide | Not toxic to grasses but kills certain leafy weeds |
| Maculosin | *Alternaria alternata* | Spotted knapweed (*Centaurea maculosa*) | Pastures and rangeland in northwestern U.S. and southwestern Canada | Host specific; best candidate for development |
| Ophiobolin I | *Drechslera sorghicola* | Johnsongrass (*Sorghum halepense*) | Corn, cotton and sugarcane crops in most of the world | Not specific, but some of its relatives seem to be specific for other hosts |
| Triticones A and B | *Curvularia clavata* | Various grasses | Lawns, crops and pastures in U.S. and elsewhere | Not specific |

Bermudagrass

Common purslane

Purple nutsedge

Crowfoot grass

Spotted knapweed

Johnsongrass

**BILLBOARD erected north of Missoula, Mont., reflects the state's concern that spotted knapweed, which is unpalatable to animals, is endangering wildlife by overrunning vast fields and choking out the food supplies of elk and other wild foragers. Unfortunately, current efforts have not been able to stop the spread of the weed (*shown in flower at the right*), either in the northwestern U.S. or even around the base of this sign itself. The author hopes that maculosin, other phytotoxins or still other biocontrol agents will eventually augment the state's ongoing control attempts.**

when applied in minute doses, the compound produced characteristic black spots on knapweed leaves. More important, it hurt only spotted knapweed and no other plants tested, including two close relatives, Russian and diffuse knapweed. Here was the first plant-specific phytotoxin ever discovered.

Stierle next made synthetic maculosin, as a first step toward another part of our work: exploring the mechanism of action. Then another graduate student, Sang Ho Park, joined the project. He created a radioactively labeled version of maculosin and showed that the weed converts maculosin to at least three other products. One or more of these products may well help produce the black lesions. Other colleagues are analyzing the chemical structure of the compounds and attempting to trace their molecular interactions in the host.

At the same time, we are wrestling with the question of why only spotted knapweed is susceptible to maculosin. We have identified several maculosin-specific receptors in the plant, which suggests that binding of maculosin to one or more of these receptors may harm the plant by interfering with the normal function of these receptors in the host. We do not yet know how this finding and the discovery of Park's three metabolites fit together. Nor do we know whether maculosin or a derivative could serve as a viable biological herbicide. We are now producing altered forms of maculosin and trying to determine whether any of them are better candidates for field testing than the original itself.

Currently the cost of spraying huge tracts of spotted knapweed with traditional herbicides is prohibitive, in part because repeated applications are needed to prevent the return of these hardy weeds. Perhaps maculosin or a still more potent synthetic variation will one day help solve the knapweed problem more economically. At the least, the compound's discovery raises the hope that other phytotoxins specific or selective for the world's many troubling weeds can be found and pressed into service—either directly or as structural models for new, environmentally compatible weed-fighting compounds.

The study of phytotoxins may lead elsewhere as well. The research is beginning to reveal new information about how fungi cause diseases of plants. By interfering with the normal activities of plants, phytotoxins are helping uncover details of plant physiology. In the future, some may even prove valuable as remedies for human disease.

For many years, investigations into the biocontrol of weeds was relatively limited and disorganized. Now, however, progress in phytotoxin research and successful field studies of other promising agents are drawing more scientists and more funding to the area. As the work continues, the strategy of "integrated" weed management—combining biological measures with modified cultivation practices and reduced use of chemical herbicides—should become increasingly common. The approach requires more effort and thought than is now generally demanded. Yet I suspect many people will decide the exertion is well worth the trouble if it yields safer food and water supplies and a cleaner environment.

FURTHER READING

BIOLOGICAL CONTROL OF WEEDS WITH PLANT PATHOGENS. Edited by Raghavan Charudattan and H. Lynn Walker. John Wiley and Sons, 1982.

POTENTIAL NEW HERBICIDES—PHYTOTOXINS FROM PLANT PATHOGENS. Doug Kenfield, Greg Bunkers, Gary A. Strobel and Fumio Sugawara in *Weed Technology,* Vol. 2, Issue 4, pages 519–524; October 1988.

MACULOSIN, A HOST-SPECIFIC PHYTOTOXIN FOR SPOTTED KNAPWEED FROM *ALTERNARIA ALTERNATA*. Andrea C. Stierle, John H. Cardellina II and Gary A. Strobel in *Proceedings of the National Academy of Sciences,* Vol. 85, pages 8008–8011; November 1988.

MICROBES AND MICROBIAL PRODUCTS AS HERBICIDES. Edited by Robert E. Hoagland. Washington, D.C., American Chemical Society, 1990.

RELEASE OF FUNGI: LARGE-SCALE USE OF FUNGI AS BIOLOGICAL WEED CONTROL AGENTS. R. Charudattan in *Risk Assessment in Agricultural Biotechnology: Proceedings of the International Conference.* Publication No. 1928. University of California, Division of Agriculture and Natural Resources, 1990.

# SCIENCE/SCOPE®

High-power ion thruster technology may provide propulsion for orbit transfer applications. The technology, under development by Hughes Aircraft Company for the National Aeronautics and Space Administration (NASA), will emphasize xenon ion thrusters in the five to ten kilowatt power range. A near-term application of this technology is electrically-powered orbit transfer vehicles that could be used for possible U.S. Air Force and Strategic Defense Initiative applications. Future technology development will concentrate on the multi-kilowatt to one megawatt power range for potential interplanetary applications.

State-of-the-art air defense systems built by Hughes protect more than one billion of the free world's population. The Air Defense Ground Environment (ADGE) systems, designed by Hughes for 23 nations, network operations centers, ground-based and airborne sensors, surface-to-air missile bases, and air bases into real-time command and control systems. ADGE systems identify all aircraft approaching their nation's borders, display the aircraft's altitude, speed, and course, and electronically interrogate the aircraft to determine its identity. Future ADGE systems will include a new distributed architecture that will allow them to use more mobile and transportable elements, as well as off-the-shelf commercial computers, for more cost-effective operation.

An improved sight stabilization system will significantly increase first-round hit probability for tank gunners. The two-axis stabilized head mirror for the U.S. Army's M1A2 Abrams tank is currently under development at Hughes. Current M1 tanks are equipped with a single-axis stabilized head mirror, which limits the gunner's ability to accurately sight and fire on moving targets when the tank is also moving. The new system is part of the Army's planned improvements for the M1A2. Hughes also produces the laser rangefinder and thermal imaging system for the current M1 tank.

An autonomous air vehicle will use artificial intelligence to perform where piloted aircraft might not survive. The Smart Weapon concept, being developed by the U.S. Defense Advanced Research Projects Agency, will execute missions into heavily defended enemy territory by using sophisticated computing technology and multi-spectral vision to replace a pilot's brain and eyes. A hierarchy of expert systems will fly the Smart Weapon from launch, identify enemy targets and defenses, and employ a "launch-and-leave" capability that fires intelligent munitions against selected targets and then returns to base. Hughes is developing these expert systems, utilizing neural networks and multi-spectral sensors.

Hughes' Microwave Products and Electron Dynamics Divisions are looking for qualified engineers in many areas. Be a part of the next generation's technological breakthroughs, working with advanced technologies such as: microwave and millimeter-wave communications, GaAs solid-state circuitry, fiber optics, image processing equipment, travelling-wave tubes, and battery design. Send your resume to Employment, Dept. S3, P.O. Box 2999, Torrance, CA 90509. Proof of U.S. Citizenship may be required. Equal Opportunity Employer.

For more information write to: P.O. Box 45068, Los Angeles, CA 90045-0068

**HUGHES**

Subsidiary of GM Hughes Electronics

© 1991 Hughes Aircraft Company

# Copper-Alloy Metallurgy in Ancient Peru

*More than 1,000 years ago civilization in South America was forged out of copper alloys. Recent excavations have revealed many aspects of copper-alloy metallurgy, from mining to smelting to metalworking*

by Izumi Shimada and John F. Merkel

From 1533 to 1534 conquistadores plundered an estimated 10 metric tons of 22-carat gold and 70 tons of fine silver from the Inca cities of Cajamarca and Cuzco. Over the past 460 years, grave looters in Peru have despoiled tens of thousands of pre-Hispanic tombs in search of gold and silver objects. In the worst cases, looters melted down precious metals and discarded any items they found made from copper and other base metals.

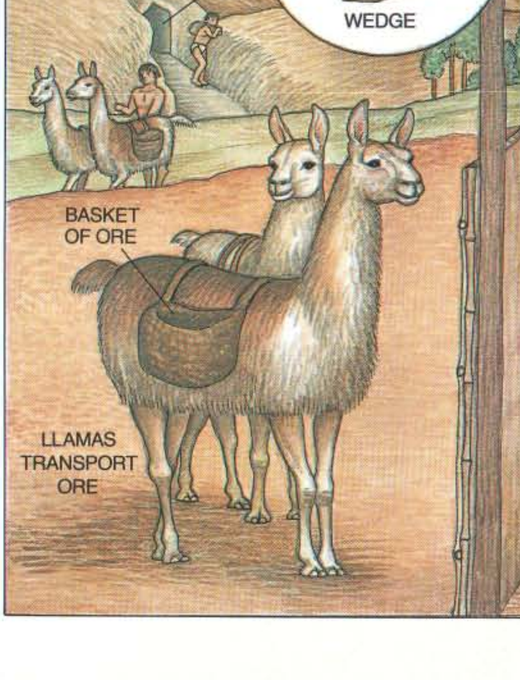
The artifacts that survived such looting were analyzed during the first seven decades of this century by such distinguished scholars as Paul Bergsoe, Junius B. Bird, Earle R. Caley, Champion H. Mathewson, N.E.H. Nordenskjöld and William C. Root. These workers documented the impressive technical achievements of native metallurgists in South America. Because they examined objects removed mainly from tombs, however, they had little information about how metals were used in the everyday lives of pre-Hispanic peoples or how they were manufactured.

Despite the gaps in the archaeological record, investigators such as Caley pieced together some details about the metallurgy of base metals in the New World. They realized that copper and copper alloys were the most widely used metals in the New World and that their production represented one of the great achievements of the peoples of ancient Peru.

IZUMI SHIMADA and JOHN F. MERKEL have collaborated for the past eight years to investigate pre-Hispanic and colonial metallurgy in northern Peru. Shimada is a research associate at the American Museum of Natural History and at the Peabody Museum of Archaeology and Ethnology at Harvard University. For the past 11 years, he has directed the Sicán Archaeological Project, which investigates pre-Hispanic Sicán culture on the north coast of Peru. Merkel is a lecturer in the department of archaeological conservation and materials science at the Institute of Archaeology, University College, London. He specializes in archaeological conservation and archaeometallurgy.

More recently, Heather N. Lechtman of the Massachusetts Institute of Technology conducted the first comprehensive survey of ancient mines and metallurgical sites in Peru. Synthesizing data from previous research as well as her own, Lechtman demonstrated that pre-Hispanic metallurgy flourished in two regions of South America: the north coast of Peru and, to the south, the high plateau from the Bolivia-Chile-Peru border to northwestern Argentina. By the second half of the first millennium B.C., artisans from both the north and south had mastered techniques for manufacturing and crafting copper. By the end of the first millennium, artisans of these regions had begun producing copper alloys in great quantities. The southern artisans emphasized alloys made of copper and tin (bronzes). Those in the north developed alloys composed of copper and arsenic. Compared with copper, both bronze and copper-arsenic alloys are much harder, stronger and easier to cast.

In 1978 one of us (Shimada) began a long-term project to understand in detail the northern metallurgical tradition. The project soon attracted geologist Alan K. Craig of Florida Atlantic University, metallurgist Ursula M. Franklin of the University of Toronto, historian Susan E. Ramirez of De Paul University, and archaeologist Stephen M. Epstein and physicist Stuart J. Fleming, both of the University of Pennsylvania. Some members of the team interviewed local residents and scrutinized Spanish colonial documents for additional data on mines, metallurgical sites and copper-alloy objects. Others surveyed pre-Hispanic mines and excavated several smelting and metalworking workshops at three metallurgical sites—all on the north coast of Peru. In 1983 one of us (Merkel) joined the project and together with other specialists analyzed metallurgical debris, ores and copper-alloy artifacts. Our team was the first to excavate pre-Hispanic copper-smelting furnaces and workshops in the New World, and at last count it had uncovered more than 50 furnaces, document-



ing how local metallurgy evolved over six centuries, from A.D. 900 to 1500.

Our fieldwork and laboratory analyses led to the first comprehensive characterization of pre-Hispanic copper-alloy metallurgy on the north coast of Peru. Based on our findings, we were able to replicate the smelting furnaces and techniques used by the ancient peoples of northern Peru. Our excavations and experiments gave us clues to the scale of copper-alloy production and its costs in terms of materials and labor. They also provided insights into the rituals associated with copper-alloy metallurgy.

As early as 500 B.C., copper metallurgy was established on the northern Peruvian coast by the Cupisnique peoples, who subsisted by farming and fishing. Some Cupisnique burials contain small copper or gold-plated copper objects. The Mochicas succeeded the Cupisnique and ruled over the north coast from A.D. 300 to 550. Known for their great temples, irrigation systems and crafts, the Mochicas greatly advanced gold and copper-based metallurgy, providing the foundation for subsequent innovations in alloying and smelting in the region.

The Mochica artisans seem to have been the first to make copper-arsenic alloys by the first few centuries A.D. They also produced a copper-gold alloy known as *tumbaga,* which was highly valued and well suited for sheet-metal work.

It was not until around A.D. 900 that the Sicáns first began large-scale production of copper-arsenic alloys. The Sicán culture, which dominated the north coast from 900 to 1100, was based on irrigation agriculture, maritime trade with Ecuador and copper-alloy production. Through the efforts of the Sicáns, the superior copper alloys replaced copper in northern Peru.

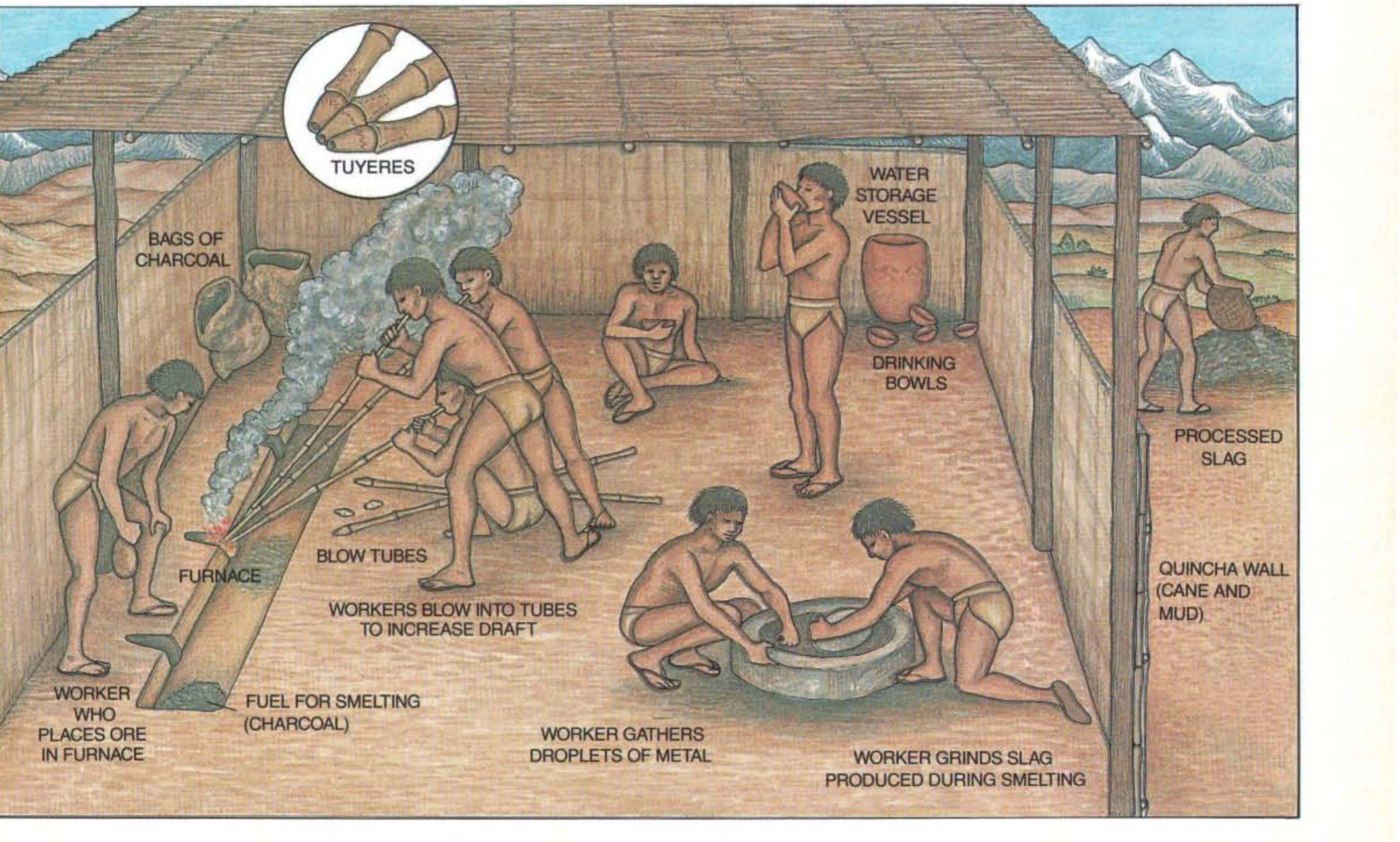
The Chimús conquered the Sicáns circa 1375 and built the largest civilization in South America before it was superseded by the Inca Empire around 1470. The Chimús transplanted the expert Sicán metalworkers to the city of Chan Chan, but metal refinement and mining remained in place. Although the Incas made an effort to spread copper-tin bronzes as the metal of the state, they apparently respected the Sicán metallurgical tradition sufficiently to allow continued production of copper-arsenic alloys.

After 1532, when Spanish explorer Francisco Pizarro and other conquistadores invaded Peru, European technology rapidly supplanted the indigenous metallurgical tradition of northern Peru.

One major site of copper-alloy production from 900 to 1532 was the Batán Grande region of the Leche River valley [*see map on next page*]. In 1978 one of us (Shimada) was drawn to this area because impressive quantities and kinds of copper-alloy objects had been looted there. The tomb of one Sicán leader, for instance, is reported to have contained copper-alloy objects weighing a total of about 500 kilograms.

Other evidence also pointed to metallurgical activity in and around Batán Grande. First, the region was strewn with a variety of equipment that had evidently been used for metalworking and smelting. Also, excavations had uncovered thick accumulations of debris that were apparently waste materials from copper production. Finally, surveys of the area had revealed ancient copper mines.

The earliest evidence of intensive smelting of copper-arsenic alloys was discovered at Huaca del Pueblo Batán Grande (which we call HPBG), located in the modern village of Batán Grande. In 1979, 1982 and 1983 one of us (Shimada) and other archaeologists excavated three deep trenches at HPBG


TUYERES
BAGS OF CHARCOAL
WATER STORAGE VESSEL
DRINKING BOWLS
PROCESSED SLAG
BLOW TUBES
FURNACE
WORKERS BLOW INTO TUBES TO INCREASE DRAFT
QUINCHA WALL (CANE AND MUD)
WORKER WHO PLACES ORE IN FURNACE
FUEL FOR SMELTING (CHARCOAL)
WORKER GATHERS DROPLETS OF METAL
WORKER GRINDS SLAG PRODUCED DURING SMELTING

and thereby uncovered four complete sets of smelting furnaces and part of a fifth. Each set consists of three to five closely spaced furnaces that were built during the Middle Sicán period (900 to 1100).

Each furnace consists of a chamber that was dug into the floor of the workshop [*see illustration on opposite page*]. The walls of the chamber are made of a clay mixture. At the top of the chamber, the walls form a narrow slot at one end and spread apart at the other, forming the mouth of the furnace. Directly beneath the narrow slot is the furnace bowl, where the fuel was burned and various mixtures of ores were smelted. A typical Middle Sicán furnace was about 30 centimeters long, 25 centimeters high and 25 centimeters wide.

The furnaces were quite well preserved. Most were intact except for some cracks in the sidewalls. The interior linings were hard and rough and in many instances covered with copper-rich materials that had corroded. Nearly all the furnaces had been partially or completely relined with mud at least twice. The relinings were apparently attempts to repair damage from repeated use.

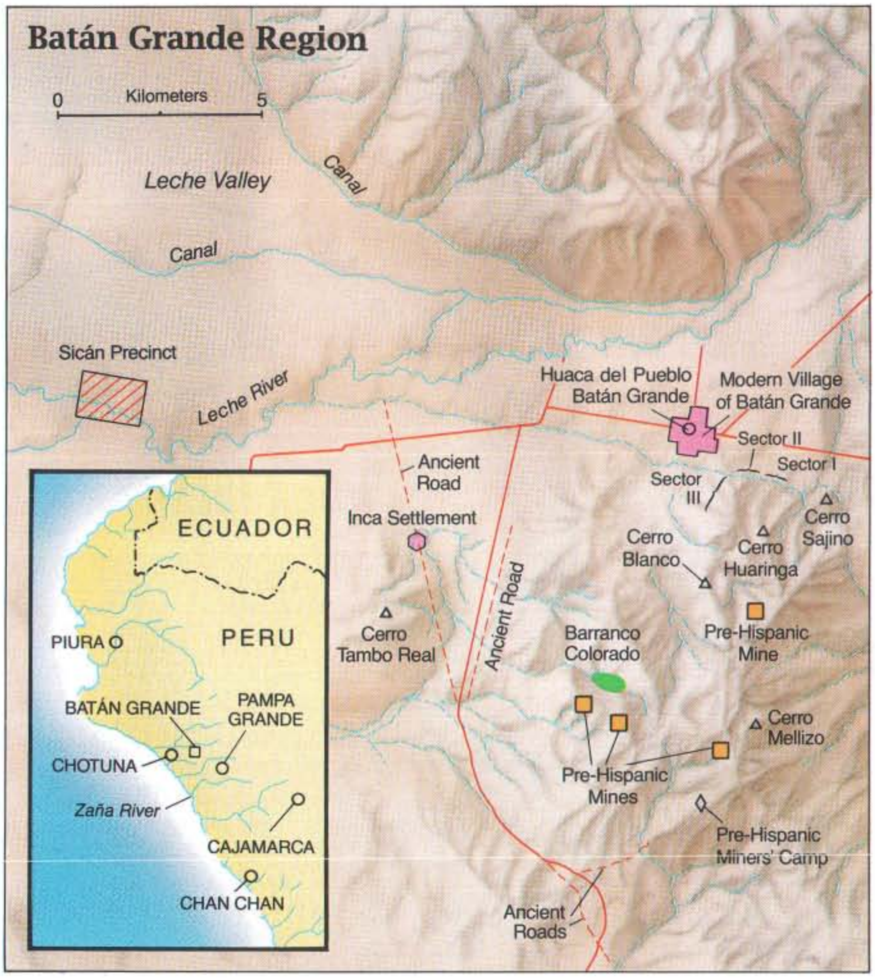
The smelting workshops at HPBG were partially demarcated by walls made from *quincha* (cane and mud). Surprisingly, the clay floors of the workshops were free of metallurgical debris, except in the immediate vicinity of the furnaces and troughs. Here the floors were littered with charcoal, ore fragments, metallurgical equipment and pieces of ceramic, or shards, some of which were covered on one side with metallurgical debris.

The excavation at HPBG also uncovered deposits of a dark, sandlike substance embedded in the workshop floors. This material proved to be slag, the fused, nonmetallic waste formed during smelting. The slag had been crushed on large anvils known as *batánes* using ovoid rocking stones called *chungos* [*see illustration on page 67*].

Our survey of nearby Cerro Huaringa documented numerous *batánes* crafted from a locally available stone called diorite. The *batánes* have highly polished central concavities about 0.75 to one meter in diameter. In fact, it is the abundance of large *batánes* that gave the name "Batán Grande" to this region.

Sicán smelting was certainly considered an important, creative act and was imbued with considerable symbolism. In other parts of the world where traditional smelting and casting are still practiced, smelting is typically preceded and followed by solemn rituals in which food, drink and other items are offered to the patron deities to ask for their assistance and give thanks.

The symbolic importance of Sicán smelting is shown perhaps most dramatically by the elaborate offerings associated with furnaces. Just before the first set of furnaces was constructed at HPBG, the Sicáns buried several fetal (or perhaps neonatal) llamas in a thick layer of organic material. This layer also covered five regularly spaced conical pits, each filled with fine gravel followed by white sand and finally capped with clay. Within a meter of each pit was a short-necked jar that perhaps once contained perishable food. Similar offerings were made when the last set of Middle Sicán furnaces was abandoned at HPBG.

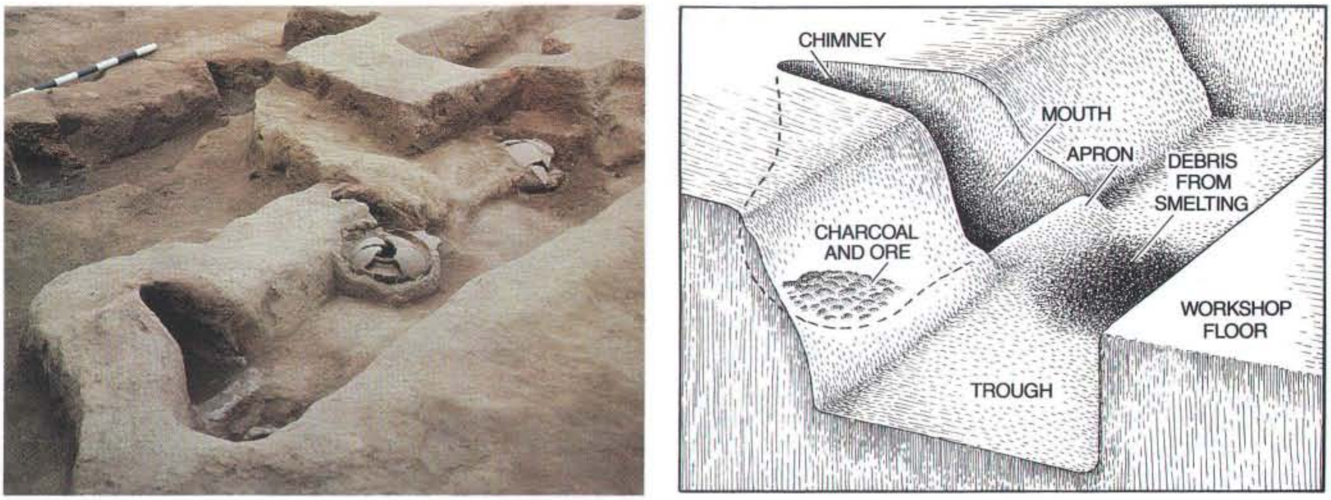

**Batán Grande Region**

Based on observations at HPBG and other sites, we were able to reconstruct the copper-arsenic-alloy smelting technology of the ancient Sicáns.

To heat the furnace, Sicán metallurgists burned charcoal at the bottom of the furnace bowl. All the charcoal found during our excavations was made from a locally available hardwood known as algarrobo (*Prosopis juliflora*). Algarrobo charcoal burns for a long time and yields much heat.

The Sicáns crushed and mixed local ores on the *batánes* to form the smelting charge. The three important components of the charge were carefully selected ores of copper, arsenic and iron. The iron served as flux, which fused with nonmetallic substances in the smelting charge so that they could be separated from the copper-arsenic metal.

The furnace held about 1.25 to 3.50 liters of charge. (We could infer the volume by examining the "waterline" of smelting debris, or slagging, on the furnace walls.) The Sicáns positioned blow tubes at the mouth of the furnace. Attached to the end of the blow tubes were ceramic tips, known as tuyeres [*see illustration on page 66*]. Many of these tips were found at HPBG and other metallurgical sites. The mouth of the furnace could probably accommodate as many as four tuyeres, judging by the shape and size of the furnace mouth.



**SMELTING FURNACES (*left*) used by the Sicáns around A.D. 900 were discovered at Huaca del Pueblo Batán Grande. Two furnaces have a votive offering placed at their mouths. The diagram at the right shows the interior of a furnace.**

By blowing through the tubes, the Sicáns accelerated the combustion of charcoal and increased the heat in the furnace. (According to historical sources, the pre-Hispanic peoples of Peru did not use any type of bellows.)

Analysis of metallurgical debris and other evidence suggest that the combustion of charcoal, enhanced by the lung-powered draft, did not produce enough heat to liquefy the charge completely. The charge was instead transformed into a thick, viscous slag containing prills, droplets of nearly pure metal several millimeters in diameter. Because the slag remained viscous throughout the smelting process, it trapped prills rather than allowing them to sink to the bottom of the furnace and form ingots.

To free the prills, the Sicáns removed the slag from the furnace and crushed it with a *batán-chungo* set. The prills were then collected by hand. We deduced that this recovery method was rather inefficient when we found that the crushed, discarded slag at HPBG contained many tiny prills (less than one millimeter in diameter). Of those prills that were collected, most were probably too small for metalworking and thus were remelted, refined and consolidated into small ingots.

The production of metals at HPBG ended around 1100. At about the same time, however, the Sicáns established smelting workshops at the northwestern base (sector III) of nearby Cerro Huaringa. By 1300 they had constructed at least 20 workshops. We found a similar number of smelting workshops on the northern slope of Cerro Sajino, three kilometers east of Cerro Huaringa. We deduced that these workshops were also established around 1100 by studying the styles of the associated ceramics and by using radioactive carbon dating techniques.

These two metallurgical sites were directly connected to a major copper mine at the base of Cerro Blanco by an ancient winding road. Although much of the pre-Hispanic mine has been destroyed over the years, we found scarred, stone hammers in nearby prospecting pits. In addition to Cerro Blanco, Sicán miners apparently exploited other mines. Within six kilometers of Cerro Blanco are the major pre-Hispanic copper mines of Barranco Colorado and Cerro Mellizo. We were able to trace parts of pre-Hispanic roads that link these mines with the Batán Grande region.

The Chimú conquest of this region around 1375 brought about significant changes in the organization of metallurgical production at Cerro Huaringa. *Quincha* smelting workshops in sector III were replaced with solid masonry constructions that all had the same architectural style. They were linked to adjacent workshops equipped with flat anvil stones on benches and vessels partially filled with charcoal, presumably for annealing. The Chimús supervised the entire area from a high platform.

When the Incas arrived around 1470, they built new smelting workshops, some *quincha* and others masonry, in sector III. They also chose to segregate the metalworking areas from the smelting workshops. Yet these changes in administration did not seem to disrupt metallurgical technology.

Most Chimú and Inca workshops were equipped with one or two *batánes* and three or four furnaces. In contrast to the Sicáns, the Chimús and Incas did not keep the workshop clean. The floors were typically littered with scraps of food, ash, charcoal bits, broken and slagged shards and fragments from tuyeres. It seems that by this time, smelting had lost much of its ceremonial significance and had become a routine, secular activity.

How did Sicán copper-alloy smelting evolve during its 600-year history at Batán Grande? The changes we have documented to date represent refinements in the basic principles and equipment applied by the Sicáns.

Overall, the design of the furnaces remained essentially the same. One notable change was a decrease in the average furnace capacity, from about 3.3 liters in Middle Sicán times to about 1.4 liters during Chimú and Inca domination. Evidence from excavations and experiments suggests that the furnace capacity was reduced in an effort to improve smelting efficiency and output, as well as furnace maintenance.

We also discovered some structural changes in the furnaces. When we examined the lining of a furnace built by the Chimús in sector III, we found that the lining was composed of a heat-resistant mixture of andesine (45 percent), mica (40 percent), quartz (15 percent) and some clay added as a binder. The constituent minerals had an average size of 0.1 millimeter, indicating that they had been carefully prepared. In contrast, earlier Sicán linings were mainly made of clay tempered with plant material and sand.

Another innovation introduced during the Chimú and Inca periods was a clay furnace constructed from two pieces. The rim of the top half of the furnace was secured to the bottom rim through a series of interlocking teeth. Besides being easier to build, two-piece construction probably facilitated repairs and relining of the furnace bowl.

We have also observed changes in the design of the tuyeres. Middle Sicán tuyeres were apparently handmade and decorated with simple geometric patterns. As the tuyeres wore down, many were repaired by smoothing the broken edges. The tuyeres of the Chimú and Inca periods were considerably larger than earlier models and were plain in appearance. Many were apparently cast in a mold, and none show signs of reuse.

Although the overall size and shape of the tuyeres changed over time, the diameter of the hole at the distal end remained constant, at about eight millimeters. The width of the furnace mouth where the tuyeres were placed also did not change much from one period to the next. Thus, the rate of airflow into the furnace probably did not vary significantly over time.

As smelting technologies improved over 600 years and the number of workshops and furnaces increased, production of copper alloys gradually intensified. In the central area of one workshop, we documented that seven sets of smelting furnaces had been built during the lifetime of the shop and that each furnace chamber had been relined two or three times. Around many workshops we discovered mounds of nearly pure ground slag as large as seven meters wide and 1.5 meters high.

For sector III of Cerro Huaringa, which was the scene of intense smelting for more than 400 years, we estimate that the amount of stratified slag is some 5,000 metric tons. As slag and other production debris piled up in the workshops during this period, the elevation of the floors rose as much as 2.5 meters.

One outstanding issue concerning pre-Hispanic production of copper-arsenic alloys is the nature of the smelting charge. Until recently, archaeologists were uncertain about the source of the arsenic and the method used for alloying it with copper. For example, the alloy could have been produced from naturally mixed ores, or alternatively, arsenic ore could have deliberately been added during copper smelting. To investigate this issue, we began analyzing the composition of prills, slag, ores and other metallurgical debris from our excavations.

Slag specimens and crushed slag found around the furnaces in Middle Sicán smelting workshops at HPBG are characteristically heterogeneous; they typically contain from 2 to 10 percent copper but are composed mostly of compounds containing iron oxide (FeO) and silicon dioxide ($SiO_2$), with some alumina ($Al_2O_3$) and calcium oxide (CaO). Such mixtures are quite similar to the slag produced by other ancient methods for smelting copper with iron-ore flux and charcoal fuel. In general, such mixtures must be heated to temperatures of 1,150 to 1,250 degrees Celsius before the slag and metal liquefy and flow freely.

The composition of copper prills encased in the furnace slag is critical for reconstructing the smelting process. The concentration of arsenic in the copper prills entrapped in HPBG furnace slag averaged about 6 percent and typically ranged from 1 to 20 percent. Some small portions of the prills contained up to 40 percent arsenic. Among the samples collected from the HPBG furnaces, we also discovered speiss, a variable mixture of copper, arsenic and iron. Speiss has a remarkable amount of arsenic (41 percent), as well as iron (56 percent) and copper (4 percent).

**TUYERES are ceramic tips that were attached to the ends of blow tubes. Sicán metallurgists shaped and decorated the tuyeres shown above.**

High-arsenic copper prills embedded in the slag, along with the presence of speiss at smelting sites, indicate that arsenic in some form was added to the charge. The arsenic could not have come from the available iron flux and copper ore, both of which usually contained less than 0.1 percent arsenic.

Since we could not readily identify specimens with low concentrations of arsenic-rich minerals at HPBG, we decided to screen more than 40 ore specimens (copper and iron) for arsenic. Two specimens contained arsenic in significant quantities, 6 and 11 percent, respectively. The specimens were different in color and density from the copper and iron ores recovered from the smelting sites, suggesting that arsenic-rich ores could be readily recognized by the Sicán metallurgists.

We also investigated ancient mines in Batán Grande, including one at Cerro Mellizo. There we found the arsenic ore known as arsenopyrite (FeAsS) and its weathering products, such as scorodite ($FeAsO_4\text{-}2H_2O$).

Much evidence at Cerro Mellizo points to the intensive pre-Hispanic mining of surface deposits of arsenic-rich minerals. Near the mine we discovered a partially preserved road bordered by stones and two sets of rooms that were built atop artificial terraces. The rooms may have been the storehouses or living quarters for miners. As indicated by ceramics found in the vicinity, the rooms were probably occupied several centuries before the Spaniards conquered Peru. In addition to Cerro Mellizo, other pre-Hispanic sites close to Batán Grande show evidence of the mining of surface deposits of copper minerals and arsenopyrite.

After studying the composition of the ores from the smelting sites and identifying at least one local mine with arsenic-rich ores, we concluded that arsenic, copper and iron ores were deliberately mixed to make the smelting charges found at HPBG and Cerro Huaringa. The low levels of arsenic in the most common copper- and iron-ore samples would not adequately account for the observed products. Although a few arsenic-rich specimens that were accidentally included in a smelting charge could produce copper-arsenic alloys, such occurrences simply would not account for the large-scale production of this alloy over some 600 years.

Although archaeological evidence and laboratory analysis revealed much information about the materials, equipment and methods used in smelting, we were still determined to figure out how the furnace actually performed and how much labor was required to operate it. In 1986 we therefore attempted to replicate the smelting techniques used by the Sicáns, Chimús and Incas. We built replicas of furnaces of various sizes and also used a 600-year-old Chimú furnace in sector III of Cerro Huaringa. The replicas were made from locally available sand, clay and rock to approximate the composition of pre-Hispanic furnaces.

We obtained copper oxide ore (31

percent copper) and hematite, an iron-containing mineral, from the mine at Cerro Blanco. After sorting the materials by hand, we crushed them using a *batán-chungo* set from sector III of Cerro Huaringa. We then mixed equal parts of copper oxide ore and hematite to make the smelting charge. For fuel we obtained small lumps of algarrobo charcoal.

After preheating the furnace for 30 minutes, we placed a small amount of charge (less than 100 grams) on top of the burning charcoal at the back of the furnace. We fit three blow tubes comfortably into the furnace mouth. As we blew air into the tubes and through the small holes in the tuyeres, we created streams of air that penetrated through the coals. The size of the stream limited the area that could be effectively heated at the front of each tuyere. To operate at maximum airflow and burning rates, two groups of three blowers alternately took shifts of roughly 10 minutes.

Using this technique, we could quickly heat a small charge to about 1,150 degrees Celsius. When we added more charge, the original material cooled and solidified into slag, and we concentrated our efforts on the unprocessed material. We could not get the entire charge to melt at any one time. We burned about two kilograms of charcoal per hour, a rate lower than we anticipated. The burning rate was the limiting time factor in the smelting process. The furnace also did not hold as much charge as we expected, because slag-coated, unburned charcoal remained at the bottom of the furnace.

At the end of some three hours of continuous blowing, during which we placed 900 grams of charge in the Chimú furnace, we created a lump of slag weighing 775 grams. The slag had nearly the same composition as those produced by the Sicáns, Chimús and Incas. We ground the slag using the *batán-chungo* set and then collected about 30 grams of copper prills.

Our experiments demonstrated that Sicán copper-alloy smelting technology was relatively primitive and very labor intensive. Even if one assumes that pre-Hispanic blowers were much more adept, each smelting would have required four to five persons working three to four hours. Given the spacing of four furnaces in the typical workshop, two furnaces could have been used simultaneously for about four smeltings per workshop in one day.

Our experiments also showed that Sicán smelting technology demanded great investment in fuel and ores. Fuel consumption in primitive copper smelting is several times greater than ore consumption. We estimate that the two dozen smelting workshops in sector III of Cerro Huaringa would have consumed more than 300 kilograms of charcoal daily.

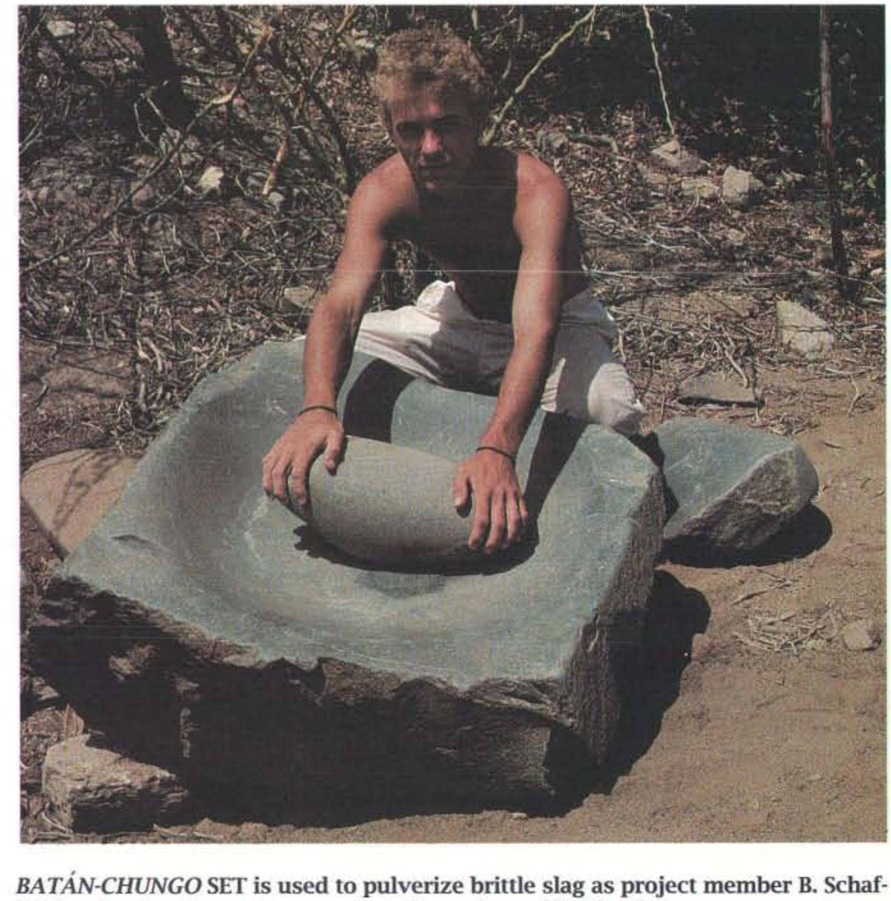
Based on estimates of slag density and furnace volume, each furnace could have held about five kilograms of molten slag. To produce that amount, one would need a charge of approximately two kilograms of copper ore (about 30 percent copper) and four to six kilograms of flux. Depending primarily on the quality of the copper-ore charges and smelting efficiency, each smelting would have produced about 0.3 to 0.6 kilogram of metallic copper.

Although these figures are rough estimates, one can now better appreciate the power commanded by the Middle Sicán leader who was buried with an estimated 500 kilograms of copper-alloy objects.

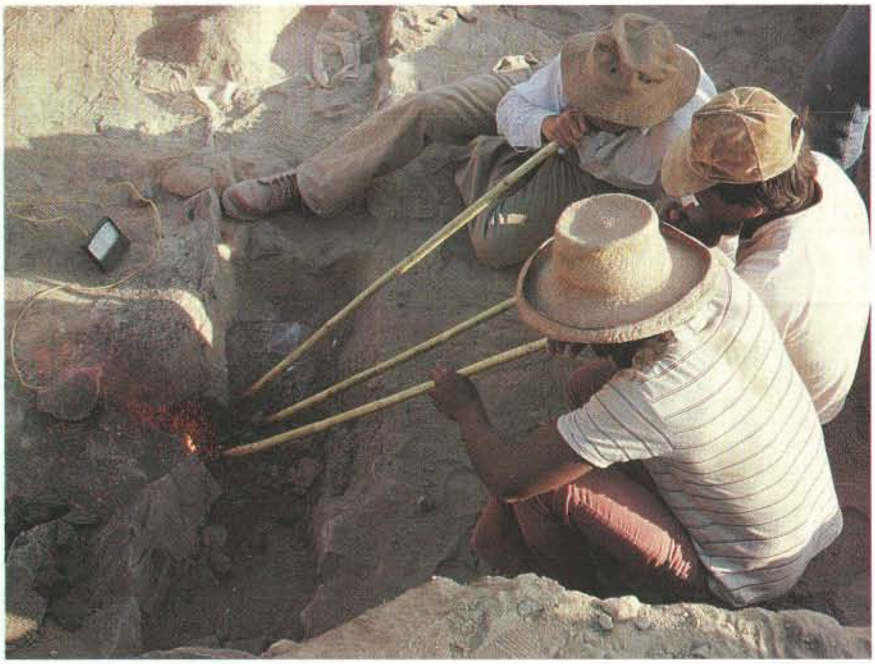
By modern standards, Sicán copper-arsenic alloy metallurgy was highly inefficient. Despite slag grinding and careful retrieval of prills, much metal was lost. The low flow of air that could be pumped into the furnaces by human lung power severely limited the amount of smelting charge that could be handled at one time. For this reason, most probably, the Sicáns had to build many smelting workshops, each with three or four furnaces.

This labor-intensive approach must be viewed, however, within its Andean context, in which access to abundant labor supply appears to have been a primary condition for productive activity. Heavy labor investment effectively countered and compensated for the intrinsic technological limitations of the primitive smelting technology.

In recent years, we and our colleagues have uncovered many copper-alloy artifacts, and we have begun to figure out some details of how ingots were transformed into finished products. From Late Sicán times until the Spanish conquest, much metalworking was performed at the northeastern base (sector I) of Cerro Huaringa. There we found evidence of hammering, annealing and finishing of objects. Scattered on the ground and on the floors of excavations were many hammer stones, anvil stones, metal fragments, slagged shards and lumps of marl—a mixture of clay, sand and calcium carbonate—that could have been used to clean and polish metal objects. Two metalwork-



***BATÁN-CHUNGO* SET is used to pulverize brittle slag as project member B. Schaffield rocks the *chungo* stone against the surface of the *batán*.**



**SMELTING EXPERIMENT performed in 1986 attempted to duplicate the techniques used in ancient Peru. Workers first placed charcoal and a mixture of ores in a 600-year-old furnace found at Cerro Huaringa. To generate a draft over the charcoal, they blew through tubes made out of cane and tipped with replicas of tuyeres.**

ing techniques seemed to have been applied most often: ingots were either hammered into sheets or cast into finished objects.

We also discovered a series of large urns that had been inverted and half-buried in a floor. Their bottoms had been carefully removed, and the urns were partially filled with charcoal bits. Urns arranged in a similar manner were also found at HPBG in a level overlying the Middle Sicán furnaces. We suspect that they could have held hot coals for annealing.

Undoubtedly, a good portion of the ingots produced in sector III of Cerro Huaringa came to sector I for further processing. In fact, sector I was connected to sector III by a major road midway up the northern slope of Cerro Huaringa. Furthermore, it is probably no coincidence that sector I is located midway between the two documented smelting sites at Cerro Sajino and sector III of Cerro Huaringa.

During the Chimú administration, most products smelted and consolidated in sector III of Cerro Huaringa, including prills, were probably exported throughout the Chimús' northern territory. Metalworking shops built during the same period have been excavated at the ceremonial center of Chotuna and the urban Chimú capital of Chan Chan.

Earlier, during the Middle Sicán period, products from HPBG may have been sent some 13 kilometers to the Sicán precinct, the religious center, which has a dozen monumental adobe temples and extensive cemeteries. In the area surrounding these temples, we recovered many broken molds, unsuccessfully cast objects, faceted stone fragments and shards coated with slag. Some ceramic vessels were lined with clay, making them stable and resistant to heat, and the vessels were also covered with slag and partially consolidated clusters of prills. The vessels were probably crucibles that were filled with prills and heated to make ingots. It is likely that the Sicán elite in the precinct supervised the manufacturing of metal objects for ritual and funerary purposes. No evidence of smelting has been found at any of these large sites, however.

Cultural attitudes toward copper-arsenic metals themselves can be inferred from the manner and context in which the metals were used. Items made of copper alloys pervaded Middle Sicán society, including everything from utilitarian bowls to ceremonial face masks. In particular, the copper alloys were worked into I-shaped, sheet-metal objects locally known as *naipes*, which are inferred to have been primitive money. *Naipes* were portable, durable and standardized in size, shape, material and manufacturing techniques. They have so far been found in Lambayeque and the Piura region (roughly matching the extent of Middle Sicán political domination). Some elite tombs in the Sicán precinct contained literally thousands of neatly stacked *naipes* of different sizes.

Many of the copper-arsenic objects were found together with *tumbaga* artifacts in the largest tombs in which Sicán elite were interred. Unlike copper-arsenic objects, *tumbaga* ornaments seem to have been exclusively for the privileged. *Tumbaga* and other precious-metal objects had a limited distribution and bore the most explicit, detailed and complete depictions of key religious icons. They were carefully assembled from pieces of sheet metal and decorated with shell and stone inlays, spangles, feathers and paint. Indeed, they were some of the most beautiful artworks created by the Middle Sicáns.

In contrast to the tombs of Middle Sicán elite, graves found near the smelting workshops in sector III—presumably interring Chimú and Inca metallurgists—are poor with respect to the quality and quantity of grave offerings. Subfloor burials in nearby metalworking shops, however, are clearly better endowed. Overall, those engaged in the busy, dirty and physically taxing work of smelting appear to have had lower social status than did metalworkers.

Current views and theories about ancient metallurgy in northern Peru still rely too heavily on artifacts gathered from tombs. To assess the significance of alloy objects in the daily life of the Sicáns, we need to excavate more households. We also plan to investigate regional and temporal variations in copper-alloy metallurgy by studying the neighboring valleys of Zaña and Jequetepeque. We can only hope that looters have not had a chance to investigate them first.

### FURTHER READING

ANCIENT AMERICAN METALLURGY: FIVE HUNDRED YEARS OF STUDY. Warwick Bray in *The Art of Precolumbian Gold: The Jan Mitchell Collection*. Edited by Julie Jones. Weidenfeld & Nicolson, 1985.

PERCEPTION, PROCUREMENT, AND MANAGEMENT OF RESOURCES: ARCHAEOLOGICAL PERSPECTIVE. Izumi Shimada in *Andean Ecology and Civilization*. Edited by Shozo Masuda, Izumi Shimada and Craig Morris. University of Tokyo Press, 1985.

TRADITIONS AND STYLES IN CENTRAL ANDEAN METALWORKING. Heather N. Lechtman in *The Beginning of the Use of Metals and Alloys*. Edited by Robert Maddin. MIT Press, 1988.

THE ANCIENT METALLURGY OF COPPER. Edited by Beno Rothenberg. London, Institute for Archaeo-Metallurgical Studies, 1990.

LOVESTORY

LEICA CAMERA GMBH, W-6336 SOLMS · PHOTO: ELLIOTT ERWITT; „PERSONAL EXPOSURES“

MADE BY LEICA

Leica products encompass a top-performance optical chain, including observation, photography and projection. These products are world-renowned and loved for their perfect detail in technology and design, for reliable mechanical precision, concentration on essentials and highest optical performance to the limits of technical feasibility. Since their first appearance more than 75 years ago, all Leica products have been characterized by an uncompromising standard of quality.

Leica

The freedom to see.

# The Austronesian Dispersal and the Origin of Languages

*The Austronesian languages of the Pacific spread across 10,000 kilometers of coastline and sea within 1,500 years, the fastest and widest expansion of prehistoric times. Farmers led the way*

by Peter Bellwood

Long before the major population movements recorded in history, many ancestral forms of the existing families of languages had already spread across enormous areas of the earth. These families—so called because their members show evidence of common descent—bear names that describe their ethnic and geographic distributions: Indo-European, Afro-Asiatic, Sino-Tibetan, Austronesian.

Why do families of languages sprawl so widely? Why, before the first historical empires and colonial frontiers, was the world not covered in a patchwork of unrelated languages? This is the situation one would expect if early populations of modern humans and their descendants had always remained in the same regions. Language distributions suggest, however, that there is much more to human prehistory than simply staying at home.

Patterns of linguistic change in historical times suggest that these prehistoric distributions resulted from long and complex processes of demographic growth and movement. It is known, for example, that the vicissitudes of 18th-century colonialism explain why French is spoken in Quebec but not in Maine. True, history does record cases of communities that have adopted the languages of their neighbors without importing the neighbors themselves: Scots, for example, speak English rather than the Gaelic of their ancestors. But such diffusion cannot explain how some languages leap entire continents, as did Arabic, English, Spanish and Russian. Whenever a language moves that far from its homeland, emigration—planned or unplanned—has always played a major role.

If a single episode of emigration suffices to spread a language, then successive episodes would seem necessary to differentiate a single language into a family of languages. The process of expansion and differentiation has been likened to biological evolution, in which a population occupies distinct niches whence it splits into subgroups, each then serving as a nucleus for further expansion.

The initial diversification of language families, and sometimes virtually all their expansionary activity, can be correlated with the archaeological record if an important theoretical assumption is made. This is that the ancestral languages from which each family stems were originally carried by expanding agricultural populations into regions that were either empty or sparsely inhabited by groups of foragers.

In the millennia that followed the last retreat of the Pleistocene glaciers, about 10,000 years ago, the increasing warmth and humidity of the middle latitudes of the earth created the current pattern of temperate and tropical climates. For reasons still not fully understood, the inhabitants of several of these regions, particularly southwest Asia, central and southern China, the New Guinea highlands and parts of Central America and western South America, began to exploit the changing environmental conditions by developing systems of food production.

To many prehistorians, this development has long been known as the Neolithic Revolution. From a modern archaeological perspective, however, the concept of revolution as opposed to gradual change appears to be applicable only to certain regions, of which the Levant and China are perhaps the best documented. The process seems to have taken relatively longer in other parts, such as Mexico or central Africa.

However prehistorians interpret the process, all surely agree on its overall significance. Now that some humans could produce their food rather than hunt or gather it, they were in a position to increase their numbers dramatically and thus to develop more complex societies. Many of these societies doubtless expanded into new territories, carrying some or all of their economies of food production with them.

In a world occupied by foragers, those groups that first developed systematic agriculture would have been able to multiply and spread for long periods, until reaching such barriers as unsuitable environments, natural disasters or the opposition of other agriculturists. The foraging populations, for their part, would have offered little resistance. In most good agricultural areas the farmers would have easily assimilated or replaced them.

If, instead, most preexisting groups had adopted agriculture themselves, then the practice would have spread mainly by diffusion and left the world with far more linguistic diversity than

PETER BELLWOOD is reader in prehistory at the Australian National University and author of five books on the archaeology of the Pacific basin. He was born in England and educated at the University of Cambridge, where he earned his doctorate in 1980. He is secretary general of the Indo-Pacific Prehistory Association and editor of its annual bulletin. He has conducted surveys and excavations throughout southeast Asia and Polynesia, most recently in the northern Moluccas in Indonesia. This is his second article for SCIENTIFIC AMERICAN.

**HAWAIIAN SAILORS reenact the epic voyages of their Polynesian forebears, the easternmost of the Austronesian-speaking peoples. Western branches carried the languages as far as Madagascar.**

it had by the dawn of history. Foragers, in fact, had compelling socioeconomic reasons to resist committing themselves to agriculture: they would have found its seasonal scheduling foreign, its work load heavy, its crowded and sedentary manner of life oppressive. Those foraging peoples that have survived were never obliged to make such a commitment and have not, at least until recently, had to compete with agriculturists for land and resources.

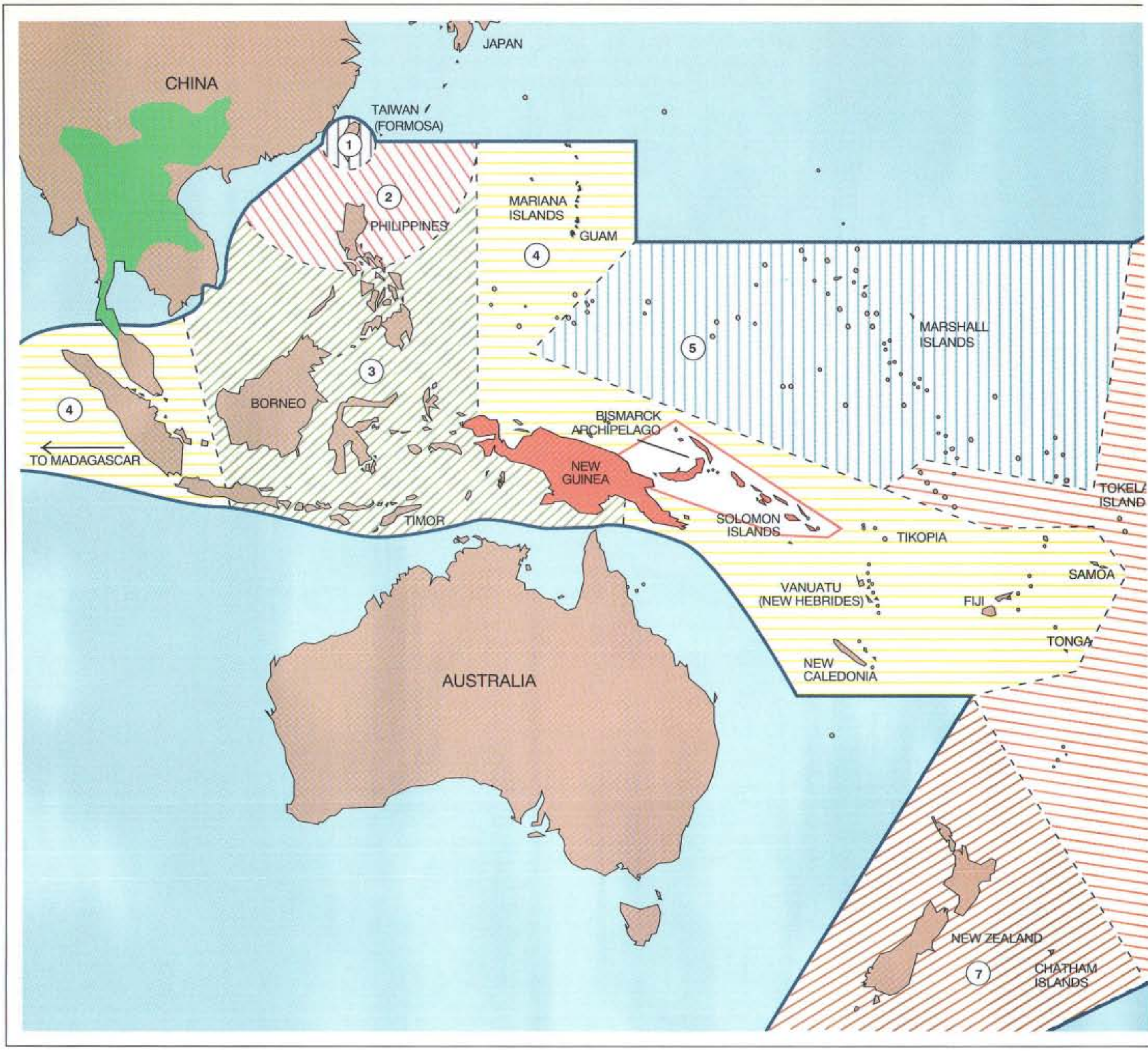
As farmers spread out, their languages spread with them. One would thus expect to find families of languages centered around regions where agriculture had been invented. These hypotheses appear to fit what is known about the development of agriculture in China and in the highlands of New Guinea: both regions developed agriculture early, and both are home to numerous language families.

Archaeological research from the Papua New Guinea highlands clearly indicates that at least 6,000 years ago people were digging ditches to control swamps, perhaps for the cultivation of taro tubers. This apparently independent development of agriculture could well have resulted in a population growth great enough to have enabled speakers of Papuan languages to expand across large areas of New Guinea, westward into Timor and Halmahera and eastward into the Solomons.

The agricultural revolutions of China seem to have occurred in two different but culturally connected foci. One—the basin of the Yellow River—saw the domestication of foxtail millet by 8,000 years ago. The other—in the Yangzi basin—saw the domestication of rice at about the same time. Both cereals, together with many less important plants, would have demanded systematic land clearance and cultivation on a seasonal basis. The quantities of their remains—often found in storage pits and habitation layers in archaeological sites—suggest that the grains rapidly attained a major dietary importance.

By 5,000 years ago, settlements of rice cultivators were in existence along the eastern coastline of China, in northern Vietnam and Thailand and possibly in northern India. In their archaeological remains are found assemblages of artifacts that leave no doubt about the overwhelming effect of the new way of life. The 7,000-year-old village of pile dwellings at Hemudu, for instance, near the southern shore of Hangzhou Bay in Zhejiang Province, has yielded pottery, matting, rope, stone adzes, wooden and bone agricultural tools, evidence for carpentry and boat building, paddles, spindle whorls for weaving, large quantities of harvested rice and the bones of domesticated pigs, dogs, chickens, cattle and water buffalo.

These are hardly the accoutrements of foragers. Indeed, their owners participated in an episode of cultural evolution that was ultimately to have repercussions over the whole of eastern Asia and the Pacific. One aftereffect may have commenced 1,000 kilometers or more south of Hemudu, where the phenomenal expansion of the speakers



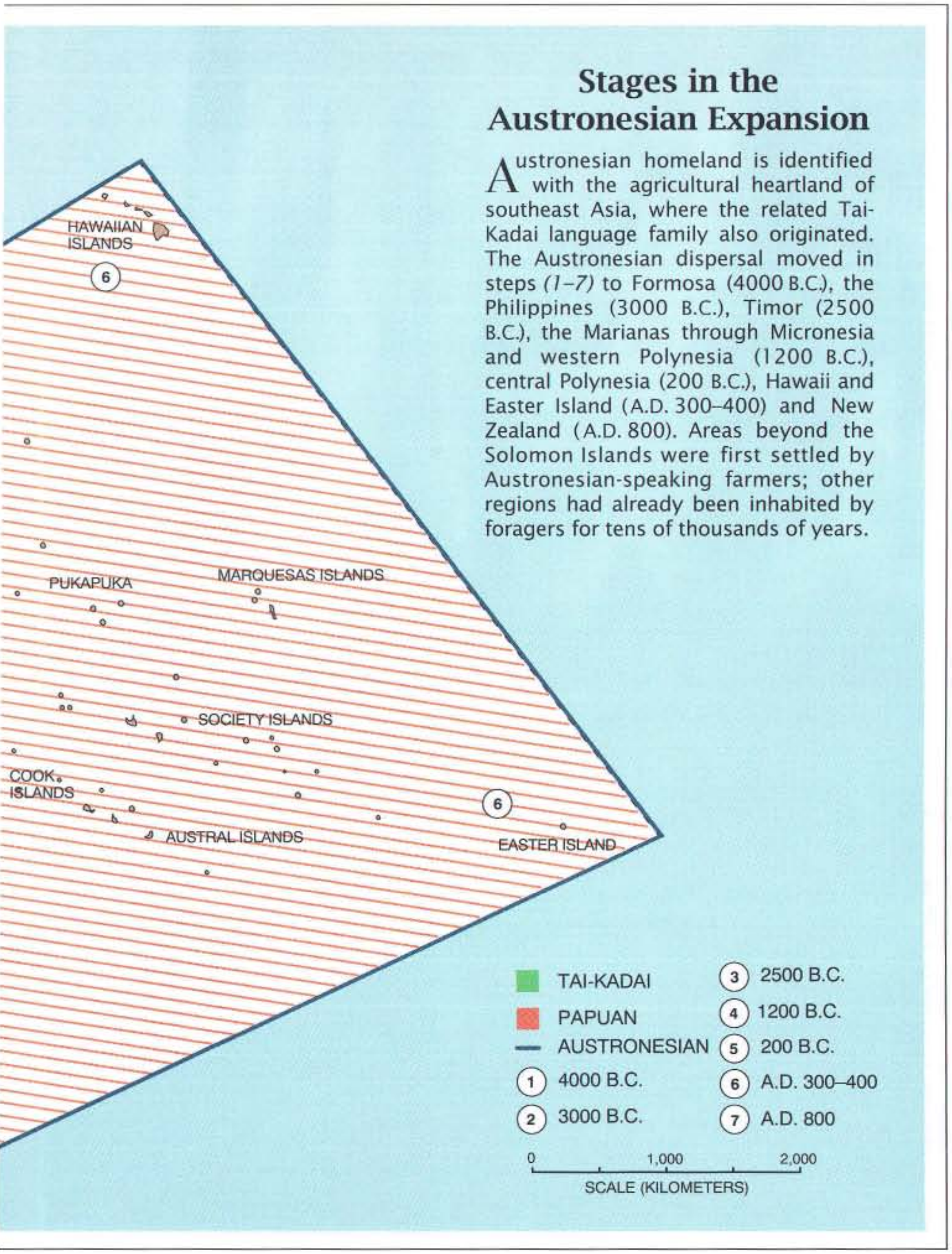
of Austronesian languages appears to have begun.

The Austronesian language family was the most widespread in the world before A.D. 1500. Today it has more than 200 million native speakers in Taiwan, inland southern Vietnam, Madagascar, Malaysia, the Philippines, Indonesia, and right through the Pacific Islands to Hawaii and Easter Island. Current archaeological evidence suggests that the family expanded first through previously settled areas, from Taiwan through to western Melanesia, and later into uninhabited territories, in Madagascar and the Pacific Islands east of the Solomons.

Evidence of the meeting of cultures in some of those lands of settlement is clearest where other families of languages break up the Austronesian distribution: in Taiwan and Vietnam, because of Chinese and Vietnamese expansion in historical times, and in western Melanesia. There the Papuan language families occupy most of New Guinea, parts of the Bismarck and Solomon archipelagoes and a few islands in eastern Indonesia. As already noted, the geographic extent of these Papuan languages probably reflects the independent development of agriculture in New Guinea, a development that seems never to have occurred in the tropical islands of southeast Asia to the west. As a result, the Austronesian languages in western Melanesia borrowed a profusion of terms and grammatical features from their Papuan neighbors, and vice versa.

Such borrowings provide evidence by which the history of language can be unraveled. Comparative linguists can, for example, infer the order in which various words were adopted from the ways in which the sounds have changed. By building a model of the changes in sound over time, linguists can test whether resemblances among languages reflect borrowing or common descent. The latter resemblances constitute the cognate set: the raw material from which linguists can reconstruct a family's ancestral speech, or protolanguage.

Reconstruction succeeds best when



### Stages in the Austronesian Expansion

Austronesian homeland is identified with the agricultural heartland of southeast Asia, where the related Tai-Kadai language family also originated. The Austronesian dispersal moved in steps *(1–7)* to Formosa (4000 B.C.), the Philippines (3000 B.C.), Timor (2500 B.C.), the Marianas through Micronesia and western Polynesia (1200 B.C.), central Polynesia (200 B.C.), Hawaii and Easter Island (A.D. 300–400) and New Zealand (A.D. 800). Areas beyond the Solomon Islands were first settled by Austronesian-speaking farmers; other regions had already been inhabited by foragers for tens of thousands of years.

there are many living languages to compare and when dead languages are well preserved in writing. Austronesian meets the former requirement; unfortunately, it lacks ancient texts outside the westerly sphere of historical Indian influence. Nevertheless, it is now possible to draw some definite conclusions, using purely linguistic evidence, concerning the homeland of the family, the directions of its subsequent spread and the cultural world of its speakers.

Commencing at the earliest time level, Paul K. Benedict, an unaffiliated U.S. linguist, hypothesizes that the Tai-Kadai language family (which includes Thai and Lao) and the Austronesian languages form a superfamily called Austro-Tai. He postulates that the superfamily had a common ancestral language or chain of languages (Proto-Austro-Tai, or PAT) once spoken on the southern Chinese mainland. Benedict has suggested a number of important lexical reconstructions for PAT, such as terms for field, wet field (for rice or taro), garden, plow, rice, sugarcane, cattle, water buffalo, ax and canoe.

If Benedict is correct—and no convincing refutation has yet been presented—then one must consider very seriously the possibility that the expansions of the Austro-Tai language family began among Neolithic rice-cultivating communities in coastal south China. The archaeological record offers ample evidence that such communities existed between 8,000 and 5,000 years ago.

Moving forward in time from Austro-Tai to its daughter, Austronesian, Robert Blust of the University of Hawaii reconstructs a family tree of successive protolanguages, beginning with Proto-Austronesian (PAN). He favors a geographic expansion that began in Taiwan (the location of the oldest Austronesian languages, including PAN), then encompassed the Philippines, Borneo and Sulawesi, and finally bifurcated, one branch moving westward to Java, the other moving eastward to Oceania via the Bismarck Archipelago.

A wealth of linguistic detail can, of course, be added to this rather bare framework, but I will restrict myself to some implications of broad historical and cultural significance. During the PAT linguistic stage, it would appear that some colonists with an agricultural economy moved across the Formosa Strait from the Chinese mainland to Taiwan. Here the archaic Austronesian language or languages developed, and from here—after several centuries—some speakers first ventured into Luzon and the rest of the Philippines. That movement divided Austronesian into its two major subgroups, Formosan and Malayo-Polynesian (MP). Before the division began, however, the PAN vocabulary indicates a culture whose economy was well suited to marginal tropical latitudes, with cultivation of rice, millet, sugarcane, the domestication of dogs and pigs and the use of canoes.

As a result of further colonizing movements through the Philippines into Borneo, Sulawesi and the Moluccas, the Malayo-Polynesian subgroup eventually separated into its several lower-order western and central-eastern branches. The divergence of central-eastern MP must have occurred in the Moluccas or Lesser Sunda islands, and eastern MP contains all the Austronesian languages of the Pacific Islands, apart from some in western parts of Micronesia. Proto-Malayo-Polynesian (PMP), which may have been spoken in the lands bordering the Sulawesi Sea, is of great interest because its reconstructed vocabulary contains tropical economic indicators absent in the earlier, more northerly PAN stage. Among these indicators are taro, breadfruit, banana, yam, sago and coconut. The PMP vocabulary also preserves terms for pottery, sailing canoes and several components of substantial timber houses.

The linguistic record in its own right informs us that Austronesian-speaking peoples—agriculturists and canoe-borne sailors—ultimately colonized all the regions occupied today by their linguistic descendants, from Madagascar to

Easter Island and from Taiwan to New Zealand. The archaeological record, our next destination, allows us to pin down this dispersal more clearly in time.

Certainly by 30,000 years ago—and perhaps as early as 50,000 years ago—sea-borne foragers had already extended their domain eastward to Sulawesi, Australia/New Guinea (then joined by a land bridge), New Ireland and the northern Solomons. To reach these places, it would have been necessary to make sea crossings of as much as 65 kilometers, the first for which human prehistory gives any record. These colonists were probably the closest direct ancestors of the modern Aborigines of Australia and the peoples of New Guinea, especially those of the interior highlands. They also contributed part of the genetic heritage of the Melanesian islanders from coastal New Guinea to Fiji (groups that later mixed with Austronesian speakers). The languages of the several Australian and Papuan families also probably descend from the languages of these first settlers.

One cannot definitively prove that the denizens of a prehistoric site or members of a prehistoric culture spoke any specific language. Here, however, circumstantial evidence supports at least the possibility of such a connection. The reconstructed PAN and PMP vocabularies represent agricultural societies that grew rice, made pots, lived in well-built timber houses and kept domesticated animals. As it happens, direct material remains of all these items survive in the archaeological record of the islands of southeast Asia and (without the rice) in the western Pacific. All the artifacts appear rather suddenly in widespread excavated sites between 6,000 and 3,500 years ago.

The Neolithic archaeological record in Taiwan began to be laid down around 6,000 years ago. It starts with archaeological assemblages of southern Chinese type, presumably carried initially by small groups of agricultural settlers across the Formosa Strait from Fujian. Characteristic artifacts found widely on the island and dated from this time onward include cord-marked pottery, polished stone adzes and reaping knives, slate spear points and baked-clay spindle whorls. There is also evidence for rice and, from pollen records, for inland forest clearance.

Between 5,000 and 4,500 years ago archaeological assemblages clearly related to these patterns spread into coastal and favorable inland regions of the Philippines, Sulawesi, northern Borneo and (together with pigs) as far southeast as Timor. Research on pollen history in the highlands of western Java and Sumatra suggests that some fairly intensive forest clearance for agriculture was under way there at least 3,000 years ago and probably earlier.

Pottery found in a site in western Sarawak in Malaysian Borneo has recently been dated to about 4,000 years ago by radiocarbon analysis of rice chaff that had been added to the clay as a tempering agent, like the straw in adobe bricks. In the equatorial latitudes of Indonesia, however, there was a shift away from rice cultivation toward a much greater dependence on the tropical fruit and tuber crops listed above for the PMP vocabulary. No cereals were ever introduced into the Pacific Islands, with the possible exception of rice in the Marianas.

By 4,000 years ago, therefore, agricultural colonists had spread in about a millennium from Taiwan to the western borders of Melanesia. The dispersal accelerated still more from Melanesia to western Polynesia, where the pottery known as the Lapita type has been found. Most of these finds exist in coastal sites and (on the smaller islands in western Melanesia) from the Admiralties in the west to Samoa in the east—a distance of about 5,000 kilometers. Last, and perhaps most awe-inspiring, was the conquest of the vast expanses of Polynesia beyond Samoa, which occurred between about 2,500 and 1,000 years ago.

### Proto–Austronesian (PAN) and Modern Derivatives

| | PAN (RECON-STRUCTED) | RUKAI (TAIWAN) | TAGALOG (PHILIP-PINES) | FIJIAN | SAMOAN | RAPANUI (EASTER ISLAND) |
|---|---|---|---|---|---|---|
| **TWO** | DUSA | DOSA | DALAWA | RUA | LUA | RUA |
| **FOUR** | SEPAT | SEPATE | ĀPAT | VĀ | FĀ | HĀ |
| **SIX** | 'ENEM | ENEME | ĀNIM | ONO | ONO | ONO |
| **BIRD** | MANUK | — | MANOK | MANUMANU | MANU | MANU |
| **EYE** | MACA | MACA | MATA | MATA | MATA | MATA |
| **EAR** | CALINGA | CALINGA | TĒNGA | DALINGA | TALINGA | TARINGA |
| **CANOE** | AWANG | AVANGE | BANGKA | WAGA | VA'A | VAKA |
| **SUGARCANE** | TEBUS | CUBUSU | TUBO | DOVU | TOLO | TOA |
| **HEAD LOUSE** | KUCUH | KOCO | KŪTO | KUTU | 'UTU | KUTU |

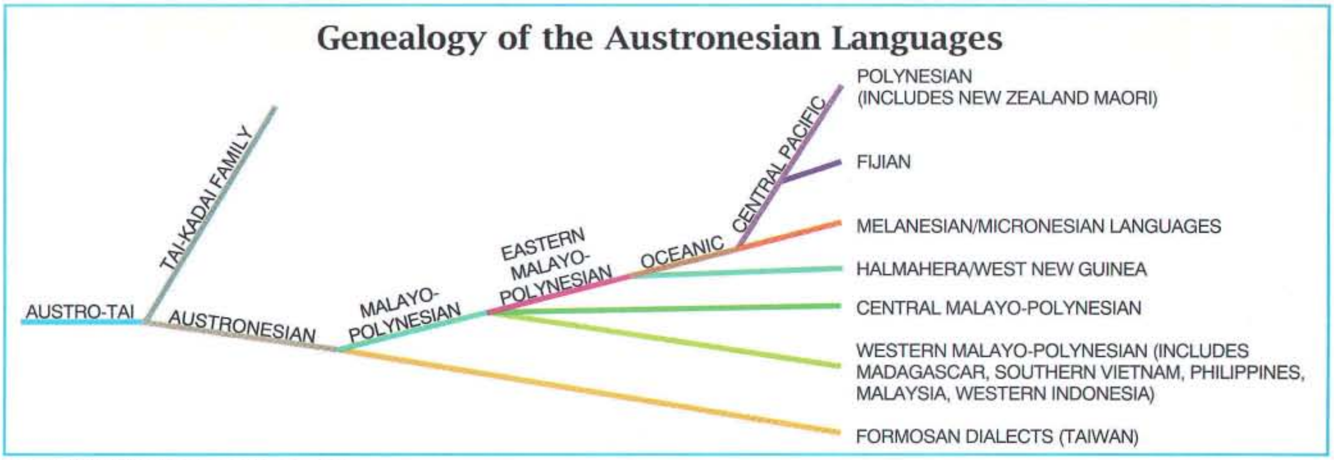
*SOURCE: Malcolm Ross, Australian National University*

**LEXICAL SIMILARITIES among the Austronesian languages indicate their descent from a common ancestor, which linguists have reconstructed. Not all roots survive in all the daughter languages: Rukai, for example, uses a noncognate term for bird.**

The main points of the archaeological record as they relate to early Austronesian dispersal and possible reasons for it can now be summarized. Austronesian-speaking agricultural colonists expanded over a period of about 1,500 years from the agricultural heartland region of southern China and Taiwan through some 10,000 kilometers of coastline and sea, from island southeast Asia to the western borders of Polynesia. The colonization proceeded more rapidly and extensively than any other in prehistoric times, bypassing the interiors of large islands in its early stages and apparently meeting with stiff cultural resistance only in western Melanesia, where archaeology indicates the existence of a prior and independent agricultural revolution.

Some idea of the remarkable navigational skills of these early Austronesians has recently been revealed by archaeology. Flakes of the sharp, glassy obsidian mined by Lapita people from sources near Talasea on the island of New Britain are now reported in sites dating to around 3,000 years ago from northern Borneo in the west to Fiji in the east, a distance of 6,500 kilometers. This probably makes Talasea obsidian the farthest distributed commodity of the whole Neolithic world. It is small wonder, then, that the descendants of these Lapita colonists were able to make the longest single voyages in human prehistory, carrying people, food plants and animals to the farthest limits of Polynesia during the first millennium A.D.

Although the Austronesian expansion was rooted in the development of agriculture, it eventually involved far more than a simple migration of a band of land-hungry farmers. Complex processes of population assimilation and interaction, adaptations of economy to differing environments and 5,000 years of innovation and enterprise all over the Austronesian world would make any simplistic view of a south Chinese or Taiwanese origin for all modern Austronesians quite unforgivable. Nevertheless, an initial expansion of linguis-


Genealogy of the Austronesian Languages
AUSTRO-TAI
TAI-KADAI FAMILY
AUSTRONESIAN
MALAYO-POLYNESIAN
EASTERN MALAYO-POLYNESIAN
OCEANIC
CENTRAL PACIFIC
POLYNESIAN (INCLUDES NEW ZEALAND MAORI)
FIJIAN
MELANESIAN/MICRONESIAN LANGUAGES
HALMAHERA/WEST NEW GUINEA
CENTRAL MALAYO-POLYNESIAN
WESTERN MALAYO-POLYNESIAN (INCLUDES MADAGASCAR, SOUTHERN VIETNAM, PHILIPPINES, MALAYSIA, WESTERN INDONESIA)
FORMOSAN DIALECTS (TAIWAN)

*SOURCES: Paul K. Benedict and Robert Blust*

tically related peoples from a Neolithic heartland forms a major theme in the Austronesian story.

Let us look more closely at this agricultural heartland, from the northern mainland of southeast Asia (including Taiwan) up to the Yellow River basin. Although the Chinese languages of the Sino-Tibetan family came to dominate the region during the past 2,000 years, it is still occupied by speakers of languages in no fewer than four other families. These are Austroasiatic (which includes Vietnamese, Khmer, the Mundaic languages of India and a number of languages of Malaya), Miao-Yao, Tai-Kadai and the Austronesian of Taiwan. With the singular exceptions of southwest Asia, western Melanesia and parts of central Africa, no other Old World territories of similar size contain as much linguistic diversity. Each exception happens to be an early nucleus of agriculture.

Equally significant is the observation that the internal diversity of these east Asian families reaches its highest degree in this region. This is what one would expect if east Asia were not only an agricultural but also a linguistic homeland. The greatest differences among the languages of a family usually derive from the most ancient splits, and the most ancient splits generally have occurred near the region in which the protolanguage was spoken.

The deepest division in the Austronesian family, for instance, is that between the Formosan and Malayo-Polynesian subgroups, which separated earliest, in Taiwan and the northern Philippines. Likewise, the internal differentiation of Tai-Kadai and Miao-Yao are most marked in southern China. For Austroasiatic and Sino-Tibetan, there is less certainty, but here, too, reconstructed vocabularies and internal divisions point strongly to an east Asian homeland.

We thus appear to be viewing a picture something like that of a budding flower, an analogy used by Andrew and Susan Sherratt of the University of Oxford's Ashmolean Museum to describe the dispersal of languages out of southwest Asia, another agricultural heartland. This region, which in its broadest sense runs from southeastern Europe and the Levant across to northwestern India, contains the presumed points of origin for the Indo-European, Caucasian, Elamo-Dravidian and possibly the Afro-Asiatic language families. It also gave rise to a number of isolated languages—such as Sumerian—that have survived only in writing. The derivation of Indo-European from southwest Asia has recently been argued by Colin Renfrew [see "The Origins of Indo-European Languages," by Colin Renfrew; SCIENTIFIC AMERICAN, October 1989].

Similar processes appear to have been at work in highland New Guinea (discussed already with respect to the Papuan language families) and also the zone of early agriculture in sub-Saharan Africa. This territory runs across the continent from the tropical west coast to Ethiopia, between about five and 15 degrees north latitude, and is the homeland of both the Nilo-Saharan and Niger-Kordofanian language families. The Niger-Kordofanian family includes the Bantoid subgroup, which expanded with agricultural colonization during the past 2,500 years throughout eastern and southern Africa, regions previously occupied by Khoisan foragers.

With these examples in mind, two hypotheses begin to take firmer shape. The first is that a heartland of early agriculture should also be a zone wherein a greater than average number of surviving language families are represented. The second is that each of the families represented should have their centers of greatest genetic diversity within the zone. These correlations need not be absolute, and in some cases (such as Turkey or much of southern China) they may be masked by language replacements that have occurred since early agricultural times. Nevertheless, both hypotheses reflect patterning that stems from the linked expansions of farmers together with their languages.

Hypotheses favoring such linked expansions have been proposed before by many archaeologists and linguists, particularly for Africa and western Asia. Even so, I do not wish to claim that early agricultural expansion explains all linguistic geography at the family level. It is obviously not relevant for traditional foraging populations such as the Australian Aborigines, and it is a fact of history that many language replacements and expansions have taken place in post-Neolithic times.

But whenever the distribution of languages cannot be explained by the interaction of early agriculturists with foragers, an analogous interaction must be sought. When one language replaces another by means of colonization, its speakers must enjoy advantages of some kind. This remains as true today as it was 10,000 years ago, when scattered bands of people first began to cultivate the earth.

### FURTHER READING

AUSTRO-THAI: LANGUAGE AND CULTURE, WITH A GLOSSARY OF ROOTS. Paul K. Benedict. HRAF Press, 1975.

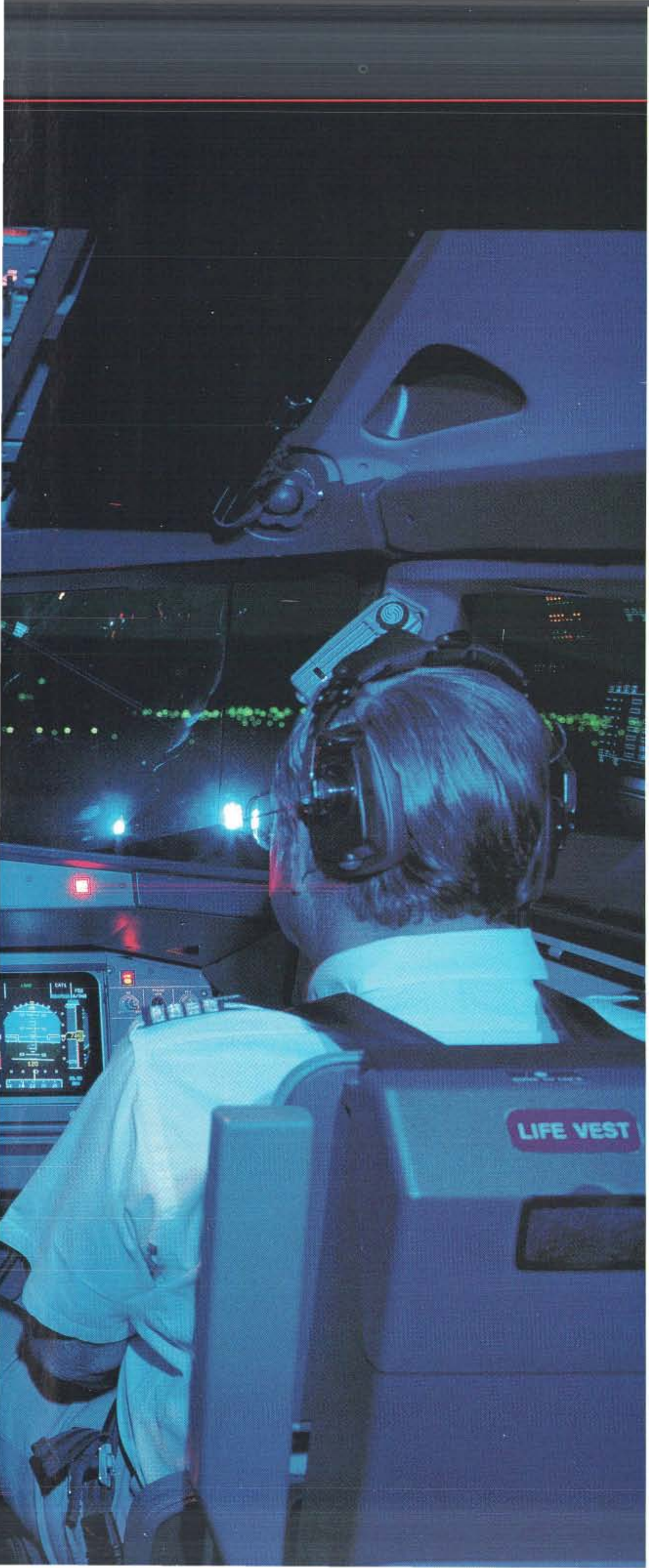
PREHISTORY OF THE INDO-MALAYSIAN ARCHIPELAGO. Peter Bellwood. Academic Press, 1985.

ARCHAEOLOGY AND LANGUAGE: THE PUZZLE OF INDO-EUROPEAN ORIGINS. Colin Renfrew. Cambridge University Press, 1987.

THE AUSTRONESIAN HOMELAND: A LINGUISTIC PERSPECTIVE. Robert Blust in *Asian Perspectives*, Vol. 26, No. 1, pages 45-67; 1988.

# ALONG FOR THE RIDE?

by Gary Stix, *staff writer*



## Computers can assume control of almost every phase of flight in the newest generation of jet aircraft.

The rear landing gear of the Boeing 737 touched down on the runway at Virginia's Wallops Island in near perfect alignment. The nose, still cocked high, eased its way onto the pavement, only a few feet from the centerline. "Not bad for a bunch of little boys from Virginia," whooped Lee Person, a test pilot for the National Aeronautics and Space Administration.

Outside, a red fox frolicked in the autumn sunlight, but neither Person nor his copilot, Cary R. Spitzer, could see it. They were staring at a bank of eight-inch-square cockpit computer screens in a windowless recess of the airplane's cabin.

It mattered little that the pilots were flying blind, because they were just along for the ride. A 16-bit microprocessor in the plane's inertial navigation system and four of a group of 16 satellites that orbit the earth every 12 hours can take credit for the landing. The NASA test plane's automatic landing and navigational computer had used signals beamed from this flotilla of satellites, the Global Positioning System, to guide the landing of the twin-engine airplane. (Two other pilots were watching through the cockpit windscreen in case the landing had to be aborted.)

Satellite-guided landings mark one more addition to a panoply of technologies that allow airplanes virtually to fly themselves. Hands-off piloting, navigation and landing of aircraft have become a routine part of commercial aviation. "Fly by wire" and "heads-up displays" (HUDs) are now common terms in aviation argot. And such tech-

*AIRBUS A320, with its highly automated cockpit and flight-control computers, has sparked debate over how much control should be handed over to the airplane and how much should remain with the pilot. The cockpit is replicated in a flight simulator in Miami.*

nologies may be supplemented by still more exotic ones: computerized expert systems may soon become assistants in the cockpit; pilots of supersonic airliners may one day maneuver by interacting with video or three-dimensional graphic representations of the world outside the airplane.

Air-traffic control will have its own complement of computer-generated expertise. The much delayed revamping of the air-traffic system may eventually result in an elaborate choreography of air- and ground-based computers. Combined wisdom from their collective circuitry can decide on who goes first and last in negotiating the bustling airspace around large metropolises.

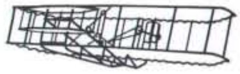
Major airframe manufacturers insist that the computers packed on board the newest generation of aircraft make those planes safer, more fuel efficient and, for the most part, easier to fly than their predecessors. With such systems, they hope to reduce the chance of human error, which is a factor in about two thirds of commercial air accidents. And with burgeoning air traffic, the number of mishaps may grow. By 2005, a major accident may plague the industry almost once every other week, Boeing officials apprehensively assert.

Where does that leave the steely-eyed aviator as automation steadily pares crew size? The redoubtable B-36 bomber of the post–World War II years carried a crew of 15, including a mechanic who would crawl into the wings to make a repair. The planes now being rolled out, even new bombers, fly with only a pilot and a copilot; the flight engineer has been largely replaced by the microprocessors in the aircraft's navigation and monitoring systems.

Aviators of all stripes, from jumbo jet pilots to executives at airlines, contend that the venerable mellow-voiced captain cannot be allowed to become merely a paid passenger or turned into a bored screen watcher. The multidisciplinary field of human factors has been enlisted to make sure that automation design keeps the pilot "in the control loop." "There's still nothing that comes close to the human being in pattern recognition and dealing with the unexpected," says John K. Lauber, a member of the National Transportation Safety Board, which investigates accidents.

## The Automating of the Airplane

| | **Pre–World War I** | **World War I to World War II** |
|---|---|---|
| **COCKPIT** | - Anemometers for airspeed<br>- Barometric altimeters<br>- Magnetic compass | - Pitot airspeed indicators (hollow metal tubes pointed into the wind)<br>- Gyroscopes and accelerometers for turning<br>- Gyroscopes for vertical and horizontal attitude |
| **FLIGHT CONTROL** | - Cables and pushrods that move flight-control surfaces on wing and tail | - Aerodynamic tabs that help to move flight surfaces<br>- Automatic pilots that maintain attitude |
| **NAVIGATION** | - Lighted beacons | - Radio beacons that broadcast Morse code to indicate that a plane has drifted off a designated path |



WRIGHT FLYER

LOCKHEED VEGA

*SOURCE: Myron Kayton*

### Before Kitty Hawk

The inexorable pace of aviation automation dates back before December 17, 1903, the day the Wright brothers first puttered aloft near Kitty Hawk, N.C. In 1891 Hiram Maxim, the inventor of the machine gun, patented a gyroscopic "stability augmentation" device that was meant to adjust the flight surfaces of his four-ton steam-powered flying machine; it succeeded in leaving the ground for a moment but never flew.

Instrumentation and control systems steadily improved the safety of aircraft and filled the cockpit with autopilots and "blind flying" instruments, such as an artificial horizon. These controls and instruments allowed the plane to fly straight and level through a dense cloud, even if the pilots removed their hands from the controls. But through it all, one thing was still clear: the crew flew the airplane.

That situation began to change in the early 1980s, with the advent in commercial aircraft of the now lowly eight- and 16-bit microprocessors. Suddenly, engineers could easily and inexpensively incorporate logic into the plane itself. And they pursued that goal with a vengeance. The first of a new generation of "smart" aircraft were Boeing's 757s and 767s and the Airbus A310, built by Airbus Industrie, a European consortium. These planes can assume the cognitive task of navigating with precision anywhere on the globe while keeping an unblinking electronic eye on hydraulic and other mechanical systems.

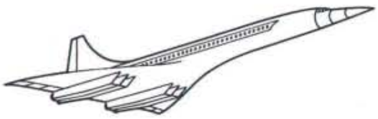
In 1988 the introduction of the Airbus A320 took aircraft automation another step further. In some respects, the flight-control computers on this highly sophisticated aircraft tell the pilot how to fly the airplane. When a pilot pushes the A320's control stick to either side, the computers let the plane bank left or right, but only so far. In effect, the software spins an electronic cocoon that stops the aircraft from exceeding its structural limitations. "This feature of the A320 sticks in the craw of most pilots," says Samuel Don Smith, a captain on Boeing 737s for Delta Airlines. Smith, who is a member of the human-performance committee of the Air Line Pilots Association, the main pilots' union in the U.S., believes that a flight crew should be able to take any action, even stressing a plane beyond its tolerances, if it is headed toward a mountain or another aircraft.

Airframe manufacturers, however, argue that the discomfort many pilots feel toward their electronic copilots is outweighed by a huge dividend in safety. Some of Boeing's newest and most automated airliners, the 757 and 767, have had one accident in nearly four million flights combined, compared with one for about every 200,000 flights for the Boeing 707, a 1950s airplane.

But even as they applaud the remarkable mechanical reliability of the modern jet airliner, the entire aviation com-

| 1945 to 1965 | 1965 to 1980 | 1980 to present |
|---|---|---|
| ▪ Analog airspeed and altitude indicators<br>▪ Radio magnetic direction indicators<br>▪ Radio altimeters<br>▪ Weather radar<br>▪ Simple alarms for fuel, temperature and landing-gear status | ▪ Mechanical flight directors that guide stick, rudder and other control movements<br>▪ Digital computers and displays that show status of hydraulic, electrical and other aircraft systems | ▪ Electronic displays with multiple levels of information<br>▪ Side-stick controllers<br>▪ Moving-map displays<br>▪ Collision-avoidance systems<br>▪ Flight management systems |
| ▪ Automatic pilots that maintain speed and direction<br>▪ Analog flight-stability computers | ▪ "Fly by oil" hydraulic system that helps to move flight surfaces<br>▪ Automatic landing systems | ▪ "Fly by wire" digital computers that send commands to flight surfaces over a network of electrical wiring |
| ▪ Omnidirectional-range beacons<br>▪ Instrument landing systems that guide a plane on its final landing approach | ▪ Inertial navigation systems that calculate a plane's position using accelerometers and gyroscopes | ▪ Satellite global positioning system<br>▪ Microwave landing systems that can guide an aircraft on a curved approach |

LOCKHEED CONSTELLATION



CONCORDE

AIRBUS A320

munity has begun to worry about what flying a plane that flies itself does to professionals who are proud of their self-control in the most trying situations. "I'm concerned that we may automate so fully that there's little or nothing left for the cabin crew to do," says James B. Busey, administrator of the Federal Aviation Administration (FAA).

Indeed, there does seem to be some justification for this concern. Last year, Airbus took the unusual step of issuing a notice warning pilots against overconfidence in flying its newest airplane, the A320. "We got the idea that some crews felt they had God on their shoulders in flying the aircraft," says Bernard Ziegler, Airbus's senior vice president of engineering and a son of one of the company's founders. "They do not. What we've built is an aircraft that is very easy to fly. But the laws of physics still apply. If you don't have enough energy to fly over an obstacle, you will hit that obstacle."

That was, unfortunately, exactly what happened to an A320 operated by Indian Airlines on February 14, 1990, when it made what is known euphemistically as a controlled flight into terrain. The crash, in which 92 people were killed, happened during a landing approach with the engines at idle, a setting for the plane's automatic throttle that is used at higher altitudes for making a descent but is never supposed to be engaged on landing. The accident followed by nearly two years another low-speed, low-altitude accident at a French air show in which three people died.

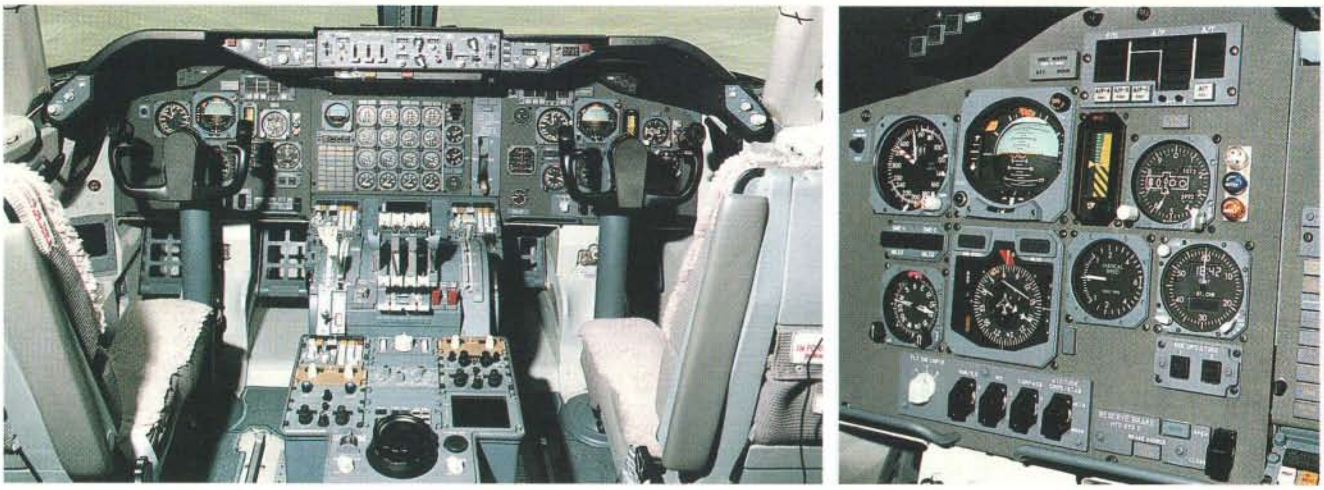
Following the Indian crash, Airbus changed software in the automatic throttle to prevent the plane from falling below a minimum authorized speed. The company also holds overconfidence prevention classes every day at the Airbus Training Center on the outskirts of the Miami Airport from five in the morning until 2 A.M. the next day.

Training against excessive reliance on aircraft systems is part of the schooling of all new A320 pilots in the center's flight simulators, two that replicate the motion of a plane in flight and one that remains stationary. A full-motion simulator, which an instructor refers to simply as the box, looks from the outside like a luggage freight container. But once inside it is so much like flying a real aircraft that some pilots who go through simulator training receive certification to fly the A320 without ever having nosed the 150-passenger, twin-engine airliner off the ground.

This $12-million three-dimensional video game—almost a third of the cost of an actual airplane—replicates the A320's "glass cockpit." Its gleaming, colored cathode-ray tubes supplant the row after row of round electromechanical instruments found in cockpits of older planes, such as the Boeing 727, the quintessential 1960s airliner that the A320 is intended to replace. There are other differences from earlier cockpits. Reminiscent of military fighters, control sticks on consoles at the far right and left of the instrument panel replace the wheellike yoke that pilots grip on other commercial jets.

In the simulator, Jerry Wolfe, an instructor for America West and a former captain with bankrupt Braniff Airlines, is putting two other veteran pilots through their paces to become certified instructors. Wolfe sits behind his students in "Cactus 1" (the radio call sign for the Phoenix-based airline) and plays the role of both instructor and air-traffic controller.

Wolfe faces a computer control panel that allows him to program a variety of flight situations that a pilot can only dread: an engine fire on takeoff, for example. A selection on the computer's menu for wind shear simply reads "Dallas," an eerie reminder of the wind condition that caused the downing of a Delta airliner in August of 1985. "I've had more emergencies in two or three periods in this simulator than I've had in my 31 years as a pilot," Wolfe re-



**1970s ROUND-DIAL COCKPIT (*left*) on the Boeing 747-200 contains a panel with instruments (*right*) for altitude, airspeed, attitude, direction and rate of climb, among other gauges, that let a pilot "fly blind" in clouds or darkness.**

marks in the drawl of his native Texas.

During one session, somewhere in the Washington, D.C., airspace, Wolfe tells Roy Taylor, the pilot of Cactus 1, to pull the control stick all the way back. As the nose points up, the plane begins to lose its aerodynamic lift. A synthesized, mechanical voice intones: "STALL, STALL, STALL." Cactus 1 has just made an alarming excursion outside what the aircraft's manufacturer calls its flight envelope.

These programmed protections that limit a pilot's actions are possible because the A320 is the first civilian aircraft to replace the mechanical system completely with digital fly-by-wire controls. (It has a rudimentary mechanical backup.) In earlier planes, a cumbersome clutch of hydraulically boosted cables, rods and pulleys translated the pilot's stick movements—or those from the automatic pilot—into a change in position of the movable surfaces on the wings and tail that maneuver the plane.

On the A320, the pilot's stick movements are sent instead to five flight-control computers that calculate how to make minute adjustments to the ailerons or other flight surfaces. The computer relays its commands via a network of wiring to hydraulic actuators on the wings or tail, which actually move the flight surfaces.

Those computers allow the pilot to push the edge of the envelope, but the plane pushes back. The computers have what flight-control specialists call "full authority," which in the A320 prevents a pilot from exceeding the specifications for safe flying. "I had to turn off two computers to get the plane to stall," Wolfe says. That action is usually forbidden by the flight manual because it inhibits the plane's maneuverability.

Although some pilots resent this lack of veto power over the computers, Airbus engineers contend that the flight envelope allows a pilot to fly unhesitatingly to the limits of the plane's capabilities. In wind shear, a pilot can pull all the way back on the side stick to nose the plane up without causing it to lurch into a deadly stall.

In any case, by mid-decade it may be difficult to find a plane without such controls. Boeing and McDonnell Douglas are expected to incorporate similar programmed safeguards in their next aircraft, the 777 and the MD-12, although pilots will have the capability of overriding them if necessary.

Military aircraft, which began to use hybrid fly-by-wire and mechanical systems more than 20 years ago, have made computer-controlled flight a staple of airframe design. Some of the newest military planes—the air force's F-117A Stealth fighter, its B-2 Stealth bomber and its F-22 Advanced Tactical Fighter (ATF), a replacement for the F-15—use three or four flight-control processors and are also built with no mechanical backup whatsoever. If power fails and an auxiliary generator does not take over, the pilot bails out or goes down with the plane.

The flight envelopes in the computers in such inherently unstable aircraft keep planes flying under conditions that few pilots could manage. In a demonstration flight, McDonnell Douglas and NASA loaded software into the flight computers of an F-15 to simulate the partial loss of the aircraft's tail from missile or antiaircraft fire. The pilot had to struggle to maintain level flight—and felt that the plane was so unsteady that he refused to make pitch or roll maneuvers. When a "self-repairing" component of the flight-control software was activated, the computers used the remaining flight surfaces to compensate for the simulated damage to the tail. A heads-up display projected onto the canopy of the fighter showed a lighted cue in a small box. Movements of the cue, tied to a pilot's control stick, represented the boundaries within which the plane could be handled.

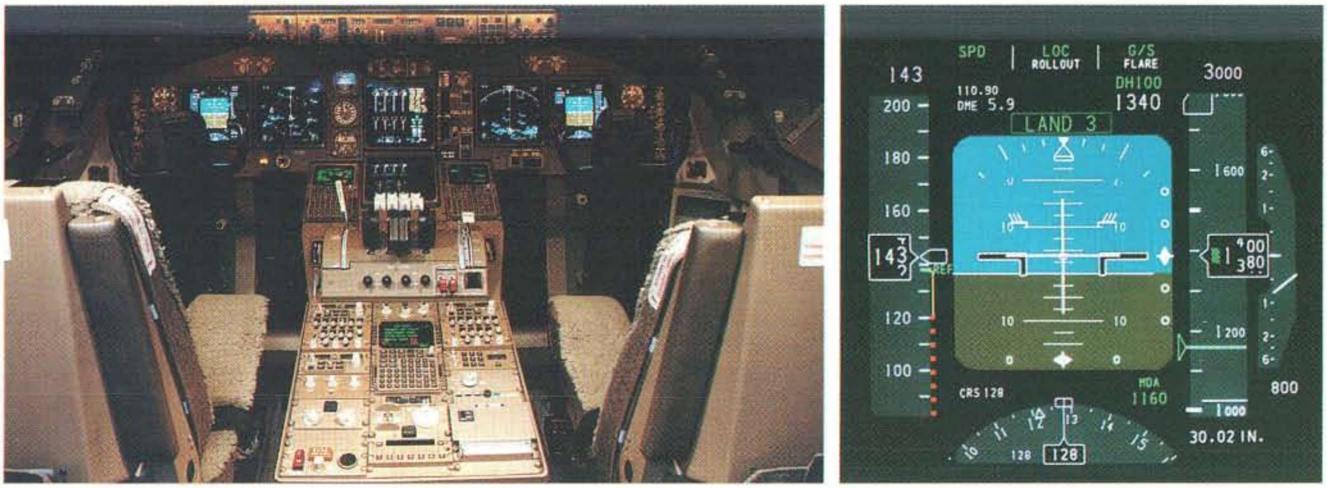
### Glass Cockpit

That modern aircraft are becoming computers with wings is most obvious in the cockpits. They are filling with enough display terminals and input devices to set up a foreign-exchange trading desk. Some pilots, in fact, refer to their cramped quarters as the office.

The glass cockpit simplifies the instrument panel, dispensing with the vast array of analog dials and gauges. By displaying information in a more concise form than a row of round dials, the cathode-ray tubes (soon to be supplanted by flat-panel, liquid-crystal displays) are supposed to simplify a pilot's "cross-check": the scan of the compass and indicators for airspeed, attitude, altitude and rate of climb. These readings are now presented on a single screen, the primary flight display.

The primary display resembles a live television program of flight. The center of the screen may be occupied by the flight director, which guides the pilot along a desired path. This instrument is bracketed on either side by tape-measure-like indicators for airspeed and altitude. Below is the cutoff upper edge of a compass dial. Wind shear and other indicators are also shown.

The role model invoked for the pilot watching these screens is that of the systems manager, a title borrowed directly from the information sciences. And that is what Arnold W. Kraby,



**1980s GLASS COCKPIT in the Boeing 747-400 (*left*) has reduced by almost two thirds the instruments found in the earlier 747-200. The "blind flying" panel has been largely replaced by the primary flight display (*right*).**

a captain for Delta Airlines, conveys to pilots making the transition from their 20-year tenure with "hand-flown" airplanes, such as the DC-9 and the Boeing 727. He has crafted a half-day course of mental preparation for training on the new airplanes.

Kraby's job is not just to reconcile Delta pilots to the high-tech gadgetry on the newest Boeing and McDonnell Douglas airplanes in the airline's fleet. He tries to give them a sense of when it is best to fly an advanced aircraft "like a 727" by overriding the system. "You get a feeling for who is going to have trouble by asking a guy whether or not he has a computer at home," Kraby says. "If he says he only uses pencil and paper, he may have problems."

The cockpit computer that provokes the most quips about a pilot's ability to type 60 words a minute is the flight management system (FMS), which made its appearance in the Boeing and Airbus airliners of the early 1980s.

The several dozen microprocessors in the FMS can ingest an entire flight plan and then assume many of the routine tasks that once absorbed the crew for most of a journey. "It gives a huge amount of information, not just about where you are now but about where you will be in the future," says Delmar M. Fadden, a senior Boeing engineer who supervises Boeing's 300-member cockpit design team, many of whom are involved with the Boeing 777.

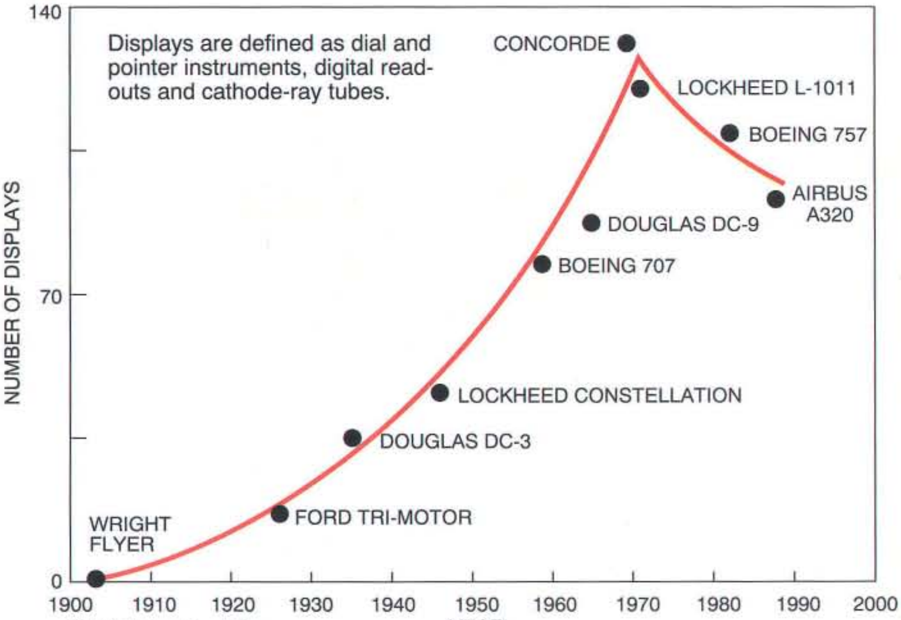
The point of entry into the FMS is a keyboard that sits on a pedestal between the pilot and first officer. There the pilot types in the route, wind forecast, altitude and the plane's weight without fuel (the system reads the fuel gauge to compute the gross weight). The FMS can calculate whether the fuel load will let the plane fly at a certain altitude, and it regulates the engine throttles to meet the most efficient fuel budget. The computer even accepts an input for a cost index, an airline-computed measure that balances fuel against the cost of paying a flight crew. If fuel is inexpensive, for example, the computer will increase airspeed in an attempt to reduce pilot and crew hours.

Reams of navigational charts, still carried on board in case of computer failure, are incorporated in the FMS software. The perspective of these digital maps, linked to the inertial navigation system, moves along with the plane. With a few adjustments, the pilot can alter the map scale from as much as 640 miles down to 10 miles—or look ahead to review a landing approach. "You tell it [the FMS] the latitude and longitude, and it knows where everything else in the world is," Kraby says.

Recently, in a 757, Kraby was headed home from Seattle to Atlanta. Just after takeoff, he received a call from air-traffic control: "Delta, you're cleared to 'RMG' "—the radio guidance beacon in Rome, Ga. Kraby typed in the letters, then hit "execute" on the FMS keyboard, and the system drew the most direct route—a "great circle" path—to the Rome radio transmitter, computing the exact distance and time of arrival. The FMS then proceeded to fly the airplane across the country by feeding navigation data to the automatic pilot.

Flight in the older Boeing 727s or DC-9s is a different experience. It con-

**Growth in the Number of Cockpit Displays**



*SOURCE: Lockheed Corp.*

sists of a series of hops between radio transmitters—very high frequency omnidirectional range beacons, or VORs. When one of these planes reaches a VOR along the route, the pilot tunes a radio to the frequency of the next transmitter using data gleaned from a paper map and flight plan.

Indeed, the FMS technology has outpaced the abilities of the air-traffic-control system. An advanced airliner can traverse a direct route, negotiate the most fuel-efficient descent into a final approach (the engines consume more fuel at lower altitudes) and calculate the time of travel almost to the second. But with the exception of sparsely traveled airspace—often in the West—such routing is seldom available because the antiquated air-traffic system cannot cope with flexible flight paths.

### Musical Runways

Earl L. Wiener, a University of Miami professor of management science and a former military pilot who now studies the ways aviators interact with their aircraft, wondered about the psychological effects of all that cockpit technology. One self-evident finding was that pilots must practice manual flying of the aircraft to retain their skills. But Wiener also discovered that cockpit automation may actually burden pilots with too much to do during the critical takeoff and landing phases of flight and not give them enough tasks to keep them alert during the monotonous middle stages. "There is a perceived loss of control on the part of some pilots," Wiener says. "They feel as if they're not really flying the airplane."

During landing, the FMS does indeed switch from being a security blanket to a demanding taskmaster. Although it does not orchestrate an actual landing, the FMS indicates on its display whether the pilot is making the correct approach, and it can automatically put the plane into the proper holding pattern if the landing has to be aborted. This scenario assumes that the air-traffic controller does not call for the plane to use a different runway, a common occurrence in the beehivelike environment around any major metropolitan airport. (The Los Angeles and Dallas airports are famous for their "musical runways.") Then, for the FMS to do its job, a first officer has to type in data furiously about the new runway.

All that urgent typing by the flight crew will be rendered unnecessary when an advanced air-traffic-control system is finally in place. With the new system, runway data will be transmitted directly to the plane's computers through a communications link that the air-traffic controller maintains with the flight being directed. The pilot will simply have to acknowledge that the plane received the new data—possibly by pushing a button on the instrument panel.

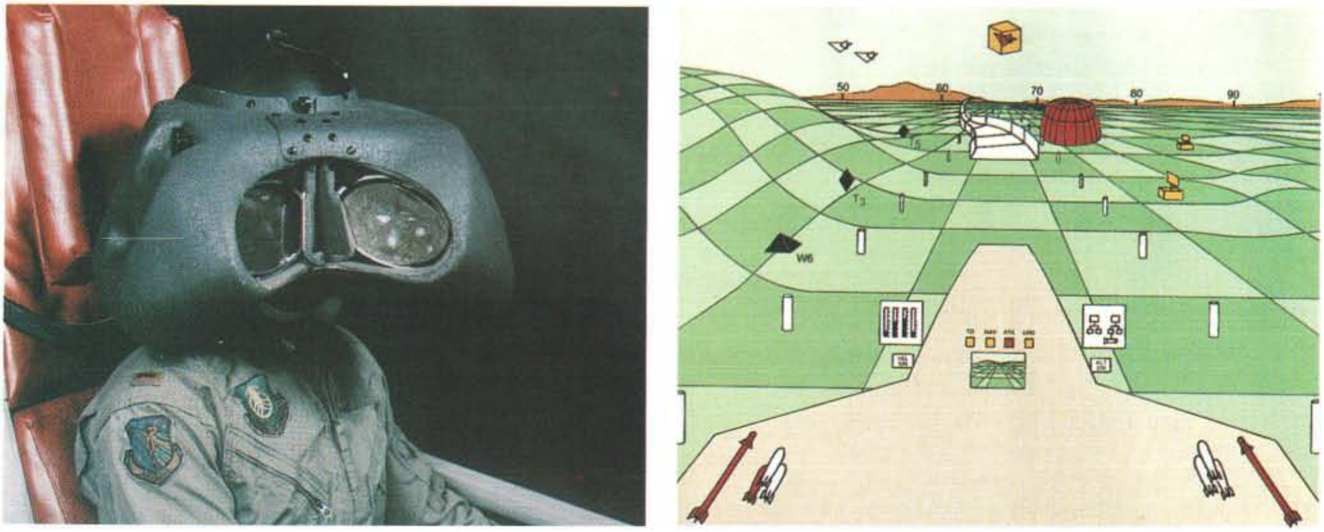
Yet some pilots fear that they will lose another cue that keeps them in contact with their surroundings: the "party-line" effect of listening to other planes talk to air-traffic controllers alerts pilots to upcoming turbulence, location of other aircraft and traffic backups. How a controller responds can reveal both competence level and work load.

Boredom and complacency, threats on the flight deck of a commercial airliner, are less of a problem for jet fighter pilots. "The cockpit of a jet fighter is the most complex machine on the face of the earth that is operated by a single human being," asserts Eugene C. Adam, a staff director for McDonnell Douglas who helped design cockpits for military aircraft ranging from the F-4 in the 1960s to one of the two competing prototypes for the ATF.

On McDonnell Douglas's F/A-18, the amount of information is as dense as antiaircraft fire. There are 675 acronyms and 177 symbols that can be displayed in four different sizes on three different cathode-ray tubes. In addition, there are 73 warning indications and 10 throttle switches; 19 controls directly under a heads-up display; and 20 controls arrayed around the three cathode-ray tubes.

The pilot's most challenging cognitive task is to make sense of the pointillistic blips from on-board radar, passive sensors detecting the emissions from enemy radar and airborne command planes relaying targeting information. These images are arrayed across several displays that the pilot must inspect individually while flying a supersonic aircraft through the chaos of battle.

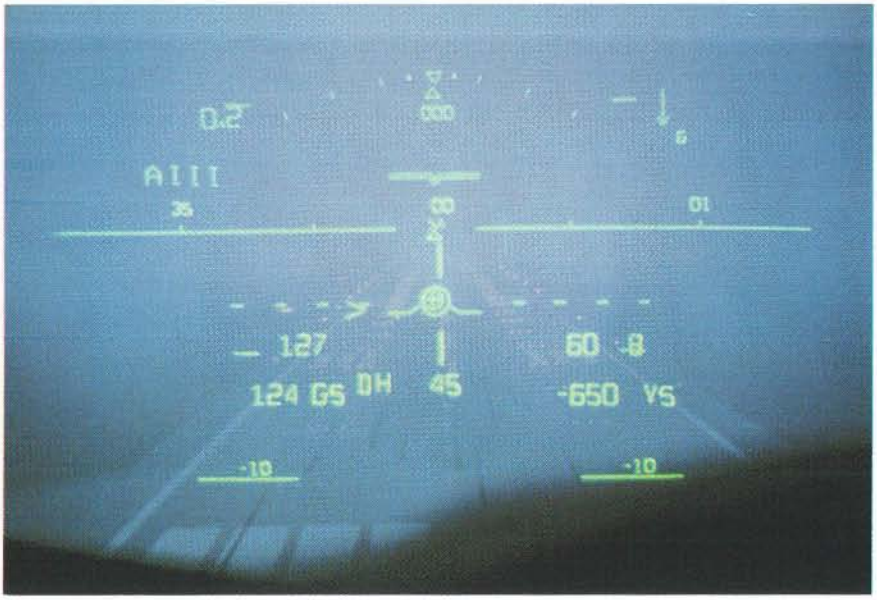
So cockpit designers for the next-generation fighter, the air force's ATF, are trying to combine various sensor inputs into a single coherent image. "Today you're given a big soup kettle called information, and you have to sort among all the data the plane is capable of displaying," says Vincent A. Devino, chief engineer for airframe and systems design for the ATF at Lockheed, which has been chosen to build the new fighter. "In the ATF," he adds, "you don't have to look at two or three different displays and assimilate what they're telling you."



**"THE BUG THAT ATE DAYTON" is an affectionate nickname for one of the three-dimensional helmet-mounted displays (*left*) developed at the Wright-Patterson Air Force Base in Dayton, Ohio. A more advanced version of such a helmet might one day let a pilot fly an actual fighter aircraft while watching a scene that resembles a video arcade game (*right*).**

With the growing ability to create realistic images on a screen, the pilot literally does not need to see outside anymore. In fact, the military's fear that low-powered lasers could be used to blind pilots has driven much of the development of advanced displays.

The Armstrong Laboratory at Wright-Patterson Air Force Base in Dayton, Ohio, is the place to get a glimpse of what might one day be possible. The laboratory has spent more than 20 years developing displays with flight data and target sights that appear on the visor of a helmet, giving a pilot the ability to aim at an enemy simply by a turn of the head. Its most notable program, the Supercockpit, was proposed in the mid-1980s as a $120-million effort to foster development of a suite of advanced cockpit technologies. The air force brass lost interest, though, and the program was never funded. The laboratory—and some of the technology envisaged for a supercockpit—survives as part of a much smaller program.



**HEADS-UP DISPLAYS projected onto a glass screen in front of the pilot use navigational data from aircraft sensors and computers to make low-visibility landings, such as the one depicted in this flight simulator.**

On a wall in the office of Wayne L. Martin, who heads the displays branch of the laboratory, hangs what looks like the picture of an arcade game. It shows a cockpit view from a cartoonlike fighter jet streaming through an unreal battle scene. A few doors away, a group of researchers have been playing with the idea of a real pilot flying an actual aircraft through such a scene.

The Wright-Patterson facility is one of the birthplaces of so-called virtual reality, three-dimensional simulations that utilize the senses of sight, sound and touch. One mammoth, insectlike helmet, used as a research tool because of its wide field of view, dates back to 1978 and looks as if it were stolen from the dressing room of the Darth Vader character in the movie *Star Wars.*

A researcher places over my head another kind of helmet, which reveals a stick-figure world from the perspective of a helicopter cockpit several hundred feet aloft. Things stir below. A point of light in the distance grows bigger and suddenly becomes a large, flickering ball. "You've just been hit," a researcher laconically announces my demise.

In another demonstration, a glove equipped with sensors that track finger movements can grasp a virtual control stick depicted on a display. A few feet away, a researcher uses only the movement of his eyes to point to different areas of a grid projected on a screen. This system can be used to manipulate objects on a display—or to aim weapons at enemy aircraft simply by looking at their image on a screen. To launch a missile, the pilot might merely push a button and utter, "Fire!"

Martin still believes that the need for such a system will become apparent by the time the technology matures, around the year 2020. "A virtual world will cost about a seventh of a conventional advanced-generation cockpit, because you eliminate the conventional controls and displays and subsequent maintenance costs," Martin says. And he is convinced that the technology will find its way into commercial aviation. A helmet equipped with a virtual reality display might be used to show a runway view to a pilot at the same time it sends an inside-the-cockpit image to an air-traffic controller.

In fact, hand-me-downs from the military could reach commercial aircraft before they are deployed on the battlefield. The NASA Langley Research Center is hard at work on video and graphics technologies that may make their way onto the supersonic transports of the next decade, which may be built without a forward cockpit view.

In one adaptation of military technology, Alaska Airlines has begun using the heads-up displays that grace fighter cockpits. The $200,000 devices, supplied by Flight Dynamics in Portland, Ore., allow a holographic image of readings for attitude, flight speed and other data to be displayed on a transparent screen between the pilot and the cockpit windscreen. A symbol on the HUD—a winged circle—guides the plane to the runway using data from the plane's navigational computers. The system can also be used on takeoffs, and it issues wind-shear alerts.

The next step may be a technology known as synthetic vision, which creates images from millimeter radar wavelengths instead of visible light. The FAA has enlisted industry and Department of Defense support for developing and testing the system to see if it can help reduce runway congestion by permitting safe landing in rain or fog.

### A Bug on the Wing?

Ironically, if FAA researchers succeed with the development of synthetic vision, they may have a difficult time getting the technology approved by the agency's regulatory side, which must certify that a technology is safe to fly. Even with a more than 800-member certification staff, the FAA has had trouble keeping up with rapid shifts in aviation technology. In the early 1980s the FAA's outdated rules did not cover the cockpit warning system for Boeing's new 757 and 767 airliners, which combined audio and visual alarms into a single system. The company succeeded in convincing regulators that the system was safe to fly.

Nearly a decade later, however, the agency has yet to update the regulatory corpus to cover this type of system. "We feel frustrated that the regulators, because of budgets and policy direction, aren't allowed to do the blue-sky thinking that's needed for new technologies," says Fadden, Boeing's senior cockpit design engineer. "So when we show up with a new concept, they don't have the experience and tools to make a timely assessment."

Rushing to judgment may be unwar-

## Where the Pilot Meets the Machine

The leading institution for research on the role human factors play in aviation lies 35 miles south of San Francisco, at the National Aeronautics and Space Administration Ames Research Center. There, in Building Number 262, an unlikely collaboration of engineers and social scientists studies the way pilots go about their jobs in a cockpit that is more a computer workstation than a driver's seat.

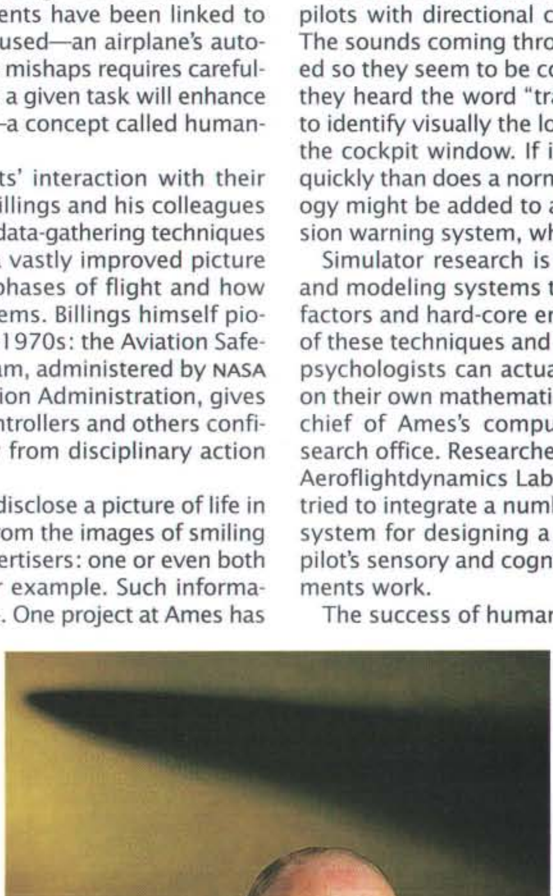
Over late afternoon tea followed by sherry, Charles E. Billings, Ames's chief scientist, complains that aircraft automation has taken on a logic of its own. The rule until now has been "it could be done, therefore it was done," says Billings, a physician by training. One result, he asserts, is that a number of accidents have been linked to the way the pilots used—or misused—an airplane's automated systems. Preventing these mishaps requires carefully assessing whether automating a given task will enhance a pilot's ability to fly an aircraft—a concept called human-centered automation.

To plumb the ways that pilots' interaction with their machines can lead to disaster, Billings and his colleagues have equipped themselves with data-gathering techniques and simulation tools that give a vastly improved picture of what happens during most phases of flight and how pilots deal with automated systems. Billings himself pioneered one such tool in the mid-1970s: the Aviation Safety Reporting System. This program, administered by NASA and funded by the Federal Aviation Administration, gives pilots, flight crews, air-traffic controllers and others confidentiality and limited immunity from disciplinary action for reporting flight incidents.

These documents sometimes disclose a picture of life in the cockpit that differs sharply from the images of smiling captains projected by airline advertisers: one or even both pilots asleep at the controls, for example. Such information can turn out to be invaluable. One project at Ames has studied the effects of sleep loss and travel across time zones. As a consequence, sleep-wake periods have been recommended for crews stopping over, and napping periods for each member of a cockpit crew, in turn, during long flights.

Researchers at Ames also study pilot behavior in flight simulators. In the late 1970s a visiting scientist at Ames, H. P. Ruffell Smith, honed the inherent realism of simulators by punctuating a quota of mechanical and weather problems with terse communications from overburdened air-traffic controllers and repeated interruptions from flight attendants.

Ruffell Smith's work turned up an unanticipated finding. Many pilot errors, he discovered, were caused not by lack of technical skills but by problems in coordination among crew members. In fact, when the hydraulic warning light flashes at 35,000 feet, the individualistic Right Stuff may be the wrong stuff. These studies have led to formal programs for pilot teamwork training that have been picked up by most major airlines and that have been emulated by nuclear reactor operators, shipboard crews and even surgical teams.



CHARLES E. BILLINGS warns of the pitfalls that may result from unbridled aircraft automation.

Ames now uses such full-mission simulations to test whether a new technology is merely an airborne pinball game or a tool that helps to give the pilot more control over the airplane. Last winter researchers turned to simulation to study a technique called three-dimensional sound.

The idea behind three-dimensional sound is to provide pilots with directional cues for responding to warnings. The sounds coming through their headphones are adjusted so they seem to be coming from a specific place. Once they heard the word "traffic," crew members were asked to identify visually the location of another aircraft through the cockpit window. If it helps pilots spot hazards more quickly than does a normal warning message, the technology might be added to a new version of an existing collision warning system, which sends out such a message.

Simulator research is being complemented by design and modeling systems that blur the line between human factors and hard-core engineering. "We've stolen enough of these techniques and put them together of late that we psychologists can actually communicate with engineers on their own mathematical terms," says E. James Hartzell, chief of Ames's computational human engineering research office. Researchers from NASA and the U.S. Army's Aeroflightdynamics Laboratory, located at Ames, have tried to integrate a number of these tools into a complete system for designing a cockpit and then simulating the pilot's sensory and cognitive responses to how the instruments work.

The success of human factors will be judged by recognition of the field's research outside NASA and a few other select institutions. Gaining acceptance for human factors is the goal of the interagency National Plan for Aviation Human Factors, which is being coordinated from the FAA by an Ames alumnus, H. Clayton Foushee. The proposal outlines a 10-year effort to focus research and to develop standards and guidelines for human factors that could be written into the safety certification rules for aircraft and other aviation systems, such as air-traffic-control workstations. The plan has been a long time in coming. "Aerospace contractors knew they should have done a better job in the human-factors area," Foushee says. "But they were not asked to, and the FAA was guilty of not specifying these things in its requirements."

ranted for a radically new design for an automatic landing system. But it has sometimes been difficult even to agree on what criteria should be used in evaluating aircraft electronics. Approving flight-critical hardware usually involves meeting a seemingly draconian requirement: the likelihood that all the computers will fail should be no more than one chance in a billion hours of flying. Fully testing such a system is impossible. There have been fewer than 20 million hours since the birth of Christ. A manufacturer can achieve that level of reliability only by adding redundant systems. Assembling three to five computers, each with a reliability rating of several thousand hours, multiplies the probability to the prescribed level.

A comparable requirement for software does not exist. The Radio Technical Commission for Aeronautics, which sponsors a joint industry-government panel that sets standards for aviation software and hardware performance, has yet to come up with a means of assessing software reliability. Software must cope with a multitude of factors difficult or impossible to quantify, including the skill of programmers as well as differences in weather, runways and other flight environments.

Critics complain that if a system cannot be built with the required margin of safety, then it should be discarded, or developers should go to the expense of providing a complete mechanical backup. "If you know you really need 10-to-the-minus-nine reliability in software and you can't get it, then you have to worry about what's meant by safety," says Bev Littlewood, a professor of software engineering at City University in London. "Maybe you shouldn't be building these systems."

The FAA relies on experience and intuition to ferret out sloppy programming. It makes selected checks of the software development process for a critical flight system. When suspect procedures are uncovered, investigators take a closer look. "If we find cancer, we generally expand the scope of the investigation," says Michael P. DeWalt, the FAA's national resource specialist for avionics software.

Bugs do, in fact, get through. Last spring, American Airlines delayed delivery of its second MD-11 from McDonnell Douglas, in part, because the cockpit displays sometimes gave incorrect readings—what the industry refers to in a new airliner as teething problems. The FAA is looking for program faults that may cause "ghost" images of another airplane to appear on the screen of a collision-avoidance device, causing pilots to take potentially dangerous evasive action in the densely packed airspace around major airports.

Mistakes in programming are compounded by the danger of pilot error. A joke told repeatedly at industry conferences puts a man and a dog in an airplane. The dog is there to bite the pilot if the man so much as tries to touch the controls; the pilot's one remaining job is to feed the dog. Many aviation veterans have heard the joke so many times that it is possible to tell those in the audience new to the industry by their laughter.

Despite the joke's staleness, there are researchers who are thinking about creating that dog. A National Research Council study group is looking at research goals to meet the needs of the nation's transportation system for the years 2010 to 2020. One topic discussed has been single-pilot commercial airliners. The first officer might be replaced by the software equivalent of the dog: a computerized expert system that would back up the remaining human.

### Pilot's Associate

Even if two pilots remain in the cockpit, there is still a need to independently check both their actions. Aviation human-factors specialists try to develop behavioral models to explain why both pilots can sometimes ignore incoming information. Forgetting to set the wing flaps and slats properly has led to two recent crashes of major commercial airliners at takeoff.

This type of research forms the basis for electronic checklists and expert systems that will diagnose engine failure, figure out whether a pilot's input to the plane's electronic systems is appropriate for a particular stage of flight or even recommend a specific course of action. Designers of these systems opt for the more euphemistic designation of "decision aiding" or "pilot's associate," to steer clear of any implication that the pilot's command may be challenged. The NASA Langley Research Center has been developing software that incorporates models of how the engine and hydraulics work in order to diagnose system failures.

The Pilot's Associate, a program sponsored by the Defense Advanced Research Projects Agency and the U.S. Air Force, is the most advanced airborne expert system. The prototype not only watches over mechanical systems but also recommends how to fly the aircraft. In one simulation of an ATF, the system automatically shut off a stuck fuel valve, mapped a course to evade a surface-to-air missile and suggested a route of return to rejoin other fighters.

Even so, the system was cast aside by the air force for the ATF. Software and hardware systems that can process these complex scenarios as they rapidly unfold are still lacking. Perhaps more important, pilots still harbor a deep antipathy to having a computer as a copilot. "Pilots want to be able to make their own decisions about whether something will kill them or not," says Lockheed's Devino, the ATF engineer.

In the race between human and dog, the human still appears to retain an edge. Yet a few people in the industry quietly acknowledge that a case can be made for letting a computer fly the airplane. By sometime in the next decade, the air-traffic-control system may need to clear planes to land at such short intervals that any last-second action by the pilot would only spell disaster.

The revamping of the air-traffic-control system calls for the development of expert systems that will help controllers schedule traffic, and cockpit computers will need to transmit automatic updates of the plane's speed, position and weight so rapidly that a pilot will barely be able to keep track of what is happening. "This linkage will make [what goes on in] the cockpit and air-traffic control almost indistinguishable," says Kevin M. Corker, a principal scientist at NASA Ames. "They'll be sharing the same information."

Could the pilotless airliner be next on design drawing boards? Not likely. Extracting a crippled airplane from a tight spot that is beyond the analytical capacity of expert software will always be the strength of the human pilot. "The human, depending on whom you talk to, is either the salvation of the system, its weak link or both," says Charles E. Billings, chief scientist at Ames. "But if you didn't have him, you would very probably have to invent him."

### FURTHER READING

AUTOMATIC FLIGHT CONTROL. E.H.J. Pallett. Blackwell Scientific Publications, 1987.

DIGITAL AVIONICS SYSTEMS. Cary R. Spitzer. Prentice Hall, 1987.

HUMAN FACTORS IN AVIATION. Edited by Earl L. Wiener and David C. Nagel. Academic Press, 1988.

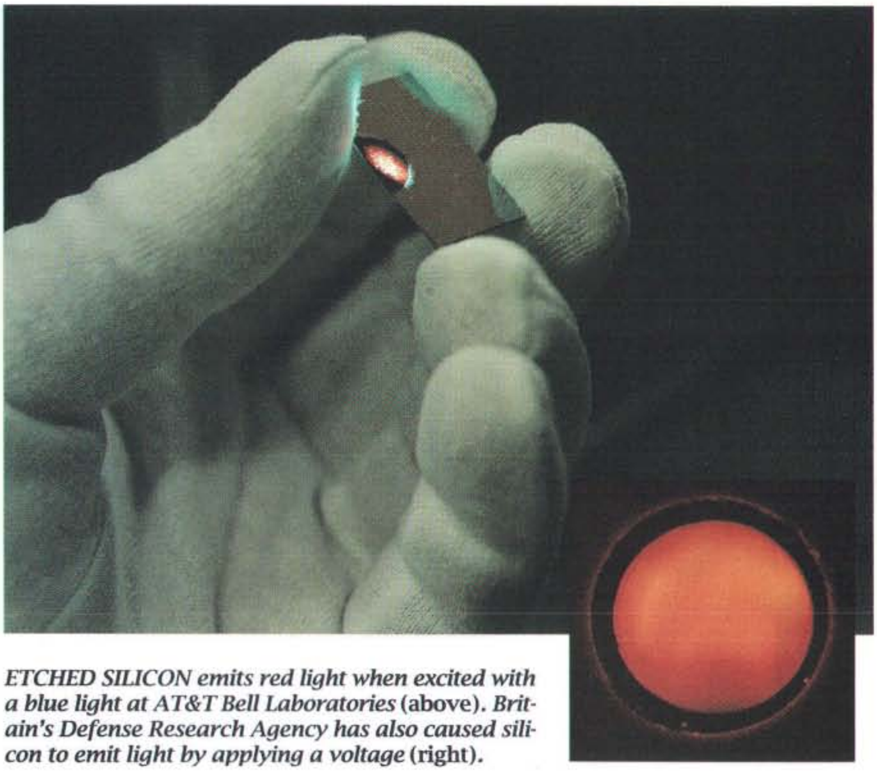
SAFE SKIES FOR TOMORROW: AVIATION SAFETY IN A COMPETITIVE ENVIRONMENT. U.S. Congress Office of Technology Assessment, 1988.

THE NATIONAL PLAN FOR AVIATION HUMAN FACTORS. U.S. Department of Transportation, Federal Aviation Administration, November 1990.

NAVIGATION: LAND, SEA, AIR AND SPACE. Edited by Myron Kayton. IEEE Press, 1990.

## Silicon Lights Up

*Researchers tease silicon into emitting light*



*ETCHED SILICON emits red light when excited with a blue light at AT&T Bell Laboratories* **(above)**. *Britain's Defense Research Agency has also caused silicon to emit light by applying a voltage* **(right)**.

Those who envision optical computers have spent the past decade or so painstakingly building light-emitting diodes from a range of finicky semiconductors. Because materials such as gallium arsenide are fragile and tricky to process, the devices are costly. A cheap optical computer, researchers said wistfully, would rely on silicon, a more manageable, abundant semiconductor. But silicon presented a vexing problem. Although it can detect and transmit light, silicon would not emit light—until now.

At the spring Materials Research Society (MRS) meeting in Anaheim, Calif., investigators from laboratories in England, France and the U.S. described how they created forests of minuscule silicon towers that emitted red, orange, yellow and green light when pumped with light of a shorter wavelength. "This is astounding behavior in a material we thought we knew everything about," says Henryk Temkin, a researcher at AT&T Bell Laboratories.

For the past 40 years, investigators have tried countless ways to tease silicon into emitting light. The problem resides largely in the electronic band-gap structure of the material. A band gap is the difference in the energy levels of the valence (lower) and conduction (upper) bands populated by electrons. Semiconductor materials that do emit light—such as gallium arsenide compounds—possess a direct band gap.

In direct band-gap materials the highest state in the valence band can be thought of as being directly below the lowest energy state of the conduction band. When electrons in the upper realm decay, they drop directly to the valence band, where they recombine with holes (positive charges), releasing photons. The energy of these photons equals the material's band gap.

Silicon, in contrast, is an indirect band-gap material; decaying electrons do not directly drop down to the valence band. To reach the appropriate level, electrons must combine with both holes and phonons, another type of energy particle. But the chance of all three particles colliding is significantly smaller—much like the reduced probability of arranging a meeting for three people instead of two, points out John C. Bean, who heads the materials science research department at Bell Labs.

To force silicon to emit light seemed to call for engineering a direct band gap. Because energy band gaps reflect the spacing of a material's crystal lattice, researchers tried stretching lattices. Investigators, including some at AT&T and IBM, succeeded in deforming the crystal structure of silicon somewhat by adding germanium—"but it didn't change silicon into a direct band-gap material," Bean concedes.

Leigh T. Canham, a researcher at the Defense Research Agency (DRA) electronics division (formerly the Royal Signals and Radar Establishment) in Great Malvern, England, tried a different strategy. From work reported in the 1950s, Canham knew he could etch tiny "worm holes," or pores, into a silicon wafer by applying hydrofluoric acid. Adding a small voltage accelerates the process. The acid chews into bumps or rough spots in the silicon.

By applying more acid to the preliminary pores, Canham gradually etched away almost 80 percent of the silicon, leaving an array of tiny towers about 50 angstroms wide. When Canham shined light from either an argon laser or even an ultraviolet source on the wafer at room temperature, the material answered with a visible red glow.

Canham believes his silicon filaments may be "quantum wires," in which the movement of electrons is confined to one dimension (up and down the wire). If so, he might have built the equivalent of a chute for electrons between the highest and lowest energy levels in the material. On the other hand, workers at the University of Grenoble in France, the first to reproduce Canham's work, have suggested that the structures may be "quantum dots," in which electrons are virtually pinned.

The investigators "changed the structure in some basic way that makes direct recombination of electrons and holes more probable," says Robin F. C. Farrow, research staff member at the IBM research division at Almaden. "We really don't understand the physics yet." Such porous silicon may turn out to be "another example where dimensionality can dramatically modify the optical properties of materials," he says.

What color silicon emits seems to depend on the amount of time the wafer has been etched: the more time spent etching a wafer, the shorter the wavelength of emitted light. As a result, Canham believes that the color of the light is a function of the thickness of the silicon filaments. "The key experiments to pin down the mechanisms have not yet been done," Bean asserts. Still, "the bottom line is that they get light out," he says, and that the results have been reproduced in other laboratories.

Ironically, given the current excitement over porous silicon, Canham's results almost escaped notice in the sea of academic papers on semiconductor materials. Although he described his findings last September in *Applied Physics Letters*, "there have been so many publications with something about getting light out of silicon that we're becoming desensitized," Bean says. "We saw the French preprint [to be published in *Surface Science* later this year] confirming the results point by point and said, 'Oh, Lord, this may do it.'" Even the organizers of the materials conference included the session on porous silicon only when Farrow mentioned reading about Canham's work in *Physics World* on an airplane flight in February.

Although the findings are provocative, these silicon light emitters are only a starting point on the road to silicon-based optical computers. To construct devices, researchers must be able to provoke light from silicon by using an electric current rather than another light. The British nonetheless claim to have done just that. At the MRS symposium, W. Yee Leong, a colleague of Canham, gave the audience a glimpse of a photograph of a silicon wafer glowing red-orange in the dark. "We're claiming we're the first to achieve stable electroluminescence in silicon at room temperature," Canham confirms.

The DRA researchers have applied for patents both on an electroluminescent device made of silicon and on a method for making silicon quantum wires. Because the DRA is a government laboratory, "we're not going to manufacture devices, but we'd like to see the technology exploited," says David A. Smith, the agency's business development manager. He hopes to find development partners and, eventually, licensees.

In spite of doubts about whether etched silicon wafers have indeed yielded electroluminescence, Bean predicts that researchers will be scrambling to find ways to achieve the results. The results published so far have opened up hitherto hidden doors in silicon research and may lead to practical, light-emitting devices. Bean predicts: "We should see some definite results on the mechanisms within three to six months." *—Elizabeth Corcoran*

## Memorable Revival

*Will ferroelectrics finally move out of the laboratory?*

In June 1955 SCIENTIFIC AMERICAN published an article on computer memories that briefly discussed a curious class of material called ferroelectrics. A ferroelectric material, the article explained, "has the property of 'remembering' the direction of an electric field applied to it." It was, the author assured readers, "another promising type of high-speed memory...in laboratory development."

More than 35 years later that description still fits. Ferroelectrics continue to promise to be the building blocks of very fast, nonvolatile memories that retain data even when power is shut off. Yet this time, ferroelectrics may finally be ready to break free of the laboratory workshop. "There is a feeling in everyone who works in this field that this is just the infancy of a discipline that once was left for dead," declares Carlos A. Paz de Araujo, a professor of electrical engineering at the University of Colorado at Colorado Springs.

The vital signs look promising. Ramtron Corporation, a start-up company in Colorado Springs co-founded by Araujo, began selling the first ferroelectric memory chips in January. Some 250 people attended this year's integrated ferroelectrics conference compared with about 50 in 1989. And several major Japanese semiconductor manufacturers have ongoing projects in ferroelectrics, according to George W. Taylor, editor of the journal *Ferroelectrics.*

In spite of the name, ferroelectrics contain no iron. Instead the term encompasses a wide range of ceramics and even some organic polymers that act somewhat like ferromagnetic compounds (which are, indeed, made of iron). Whereas ferromagnetic materials become polarized near magnetic fields, ferroelectrics become spontaneously polarized when exposed to an external electric field. In addition, some ferroelectrics may also exhibit sensitivity to light, infrared radiation and mechanical stress—properties that some researchers are also trying to exploit.

Because ferroelectrics remain polarized even when a field is removed, they can provide the makings of nonvolatile computer memories, chalking up either a "1" or a "0," depending on how electric charges are distributed in the ferroelectric. To switch the contents of a ferroelectric memory (say, to store a "0" instead of a "1"), researchers need only expose the material to another electric field of the opposite polarity.

Previous efforts to develop ferroelectric memories failed when engineers tried to fabricate devices. Often the researchers used slabs of relatively thick ferroelectric films that could not be easily controlled. Moreover, there was no driving need for extremely fast memory devices; when paired with available microprocessors, the contemporary memories performed adequately, recalls Donald R. Lampe, an advisory engineer in the advanced technology division of Westinghouse Electric in Baltimore. Yet as consumers began to demand faster switching speeds and less power consumption, Westinghouse as well as other manufacturers began to look back and say, "How about ferroelectrics?" Lampe says.

The few other nonvolatile memories have limitations. The most widely used devices—dynamic random-access memories, or DRAMs—must be continually refreshed to retain data. Military equipment that requires nonvolatile memory still typically depends on relatively slow and very expensive memories. Commercial manufacturers have turned to electronically erasable, programmable, read-only memories (EEPROMs) and more recently to so-called Flash memories.

Yet EEPROMs and Flash memories involve trade-offs. Storing, or writing, data on an EEPROM is time-consuming: it takes between one and 10 milliseconds per byte. Moreover, about 10,000 writing operations fatigue the memory. (Retrieving, or reading, data is painless and swift, however.) Flash memories, which experts describe as advanced EEPROMs, write data far more quickly—often in less than 0.1 millisecond per byte. Yet they, too, can endure only about $10^4$ write operations.

In principle, ferroelectrics can store data several orders of magnitude faster than can Flash memories. Ferroelectrics are also largely impervious to radiation, a trait essential for military or space-bound components.

There are, however, two possible drawbacks. First, reading data from a ferroelectric memory, which requires determining the polarization of the memory, often destroys the data for a few fleeting nanoseconds. (They are then quickly rewritten.) Military users consequently worry that if such chips encounter a sudden burst of radiation, data could be irrevocably lost. Other users are concerned that repeated reading and writing degrades the ferroelectric films. Although these materials can generally withstand about $10^{12}$ changes in polarization, that endurance may still fall short of the countless reads (albeit limited writes) of EEPROMs or Flash memories.

Researchers are nonetheless sanguine about ferroelectrics. "I don't think there is a problem," insists David W. Bondurant, who directs marketing at Ramtron. In Ramtron's four-kilobit FRAM, the first ferroelectric memory on the market, engineers substituted a thin layer of the ferroelectric ceramic lead zirconate titanate for the capacitors in a conventional DRAM. During routine use, the device performs like a standard DRAM, so that writing operations do not degrade the ferroelectric material. When the power is shut off or interrupted, however, the data are stored in the permanent ferroelectric memory. Bondurant says the company hopes to have larger-capacity chips available by the end of the year.

National Semiconductor Corporation in Santa Clara, Calif., plans to introduce its own 4K ferroelectric memory chip late this year. "I think we have a viable product," asserts Norman E. Abt, a principal engineer at National. "The problem is proving the reliability," he adds. Past research on ferroelectrics has focused more on the reliability of bulk materials. Because the relatively new, thin films will behave differently from bulk materials, engineers are still studying how the materials will hold up over time. "How far can we go with the technology?" Abt asks. "I've seen nothing in the physics that says larger chips won't work."

Other researchers are also hoping to read such memory chips nondestructively by taking advantage of ferroelectrics' sensitivity to light. Exposing a film to light generates a tiny current that can indicate the contents of a memory. Such devices can be blisteringly fast, reports Sarita Thakoor, a researcher at the Jet Propulsion Laboratory in Pasadena, Calif. "We have been able to get millions of repetitive readouts very fast with negligible change in the material or stored memory," she says. She intends to incorporate such devices in optical neural networks.

Rather than trying to build commercial DRAMs, Lampe and his colleagues at Westinghouse plan to exploit the fast switching properties by incorporating the ferroelectric material barium magnesium fluoride into the gate of a field-effect transistor. By using the ferroelectric to open and shut the transistor gate, as in an EEPROM, the workers say they will avoid losing data. Lampe expects to begin building experimental chips "momentarily."

A handful of other researchers are trying out ferroelectrics in electromechanical applications. Dennis L. Polla, a professor at the University of Minnesota at Minneapolis, for example, began building micromechanical devices with ferroelectrics when he realized the compounds produced higher voltages when flexed than did other piezoelectric materials. Recently he built a minuscule ferroelectric pressure sensor for improving acoustic detection in such devices as hearing aids and a micropositioning bar to control tiny movements of larger instruments.

Because his ferroelectric films lie on top of silicon substrates, Polla is confident that he will be able to integrate the micromechanical components into a conventional, silicon-based electronic circuit. "There aren't materials you can easily substitute for ferroelectrics unless you either give up using silicon substrates or cryogenically cool your devices," he adds.

Workers at Sandia National Laboratories are using ferroelectric films to construct microrefrigerators, reports Bruce C. Bunker, who supervises the electronic ceramics division. These devices, which could pump gases through tiny channels, may prove handy for cooling circuits.

Perhaps the most dangerous pitfall, some experts say, is that the boom in interest will reinflate researchers' ambitions before they have a solid grasp of the properties and characteristics of ferroelectrics. If results fall short of ideal, people may once again shy away from ferroelectrics, cautions Richard L. Wiker, an engineer at the space and strategic avionics division of Honeywell in Clearwater, Fla. "There are about 1,500 possible ferroelectric materials," he says. "We don't think the final one has been selected yet." —*Elizabeth Corcoran*

## A Chip-making Plan to Leapfrog Japan

Remember the Chip Wars? A few years ago anxiety over the faltering American semiconductor industry led the U.S. to establish a joint industry-government research consortium called Sematech to improve domestic chip-manufacturing prowess. This summer industry leaders are pulling together plans for the next phase of their comeback play: they call it "Micro Tech 2000."

Over the past 20 years or so the density of chips—say, how many transistors can fit on the surface of a chip—has quadrupled every three years. To jump ahead of the Japanese, experts argue, U.S. industry must vault over a generation and produce a static random-access memory chip that can handle one billion bits of data (a gigabit). To investigate the feasibility of such a leap, the congressionally appointed National Advisory Committee on Semiconductors (NACS) is creating a "technology road map" that will point out what the U.S. semiconductor industry must do to get from here to there.

According to James D. Meindl, provost of Rensselaer Polytechnic Institute, chips in the future will be "systems on a chip," devices that integrate a gigabit of memory with a microprocessor and other functional units. Building such chips will demand detailed computational models that can simulate every aspect of production and operation—from the manufacturing line to the final component board. "The future chip factory will have a large software content," adds William J. Spencer, chief executive officer of Sematech, based in Austin, Tex.

Developing those factories will not be cheap. According to NACS, a gigabit-chip fabrication facility will require a staggering $2 billion investment. As a result, only 25 or 27 such advanced plants will be built, Meindl predicts, "and fewer than 10 of them would be built in the U.S." Such consolidation calls for increased cooperation among U.S. chip makers, he says. The advisory committee hopes to complete its assessment of the technical steps needed to build a one-gigabit chip by late July. By the end of the year, the group will also release recommendations on precisely how the U.S. semiconductor industry might undertake such an effort.

Some have predicted that effort will become an industrial "Apollo" project. Sematech, which will be involved, is an important role model, adds Robert W. Galvin, former chairman of Motorola. In such forums, "competitors are finding it honorable and practical to cooperate with each other," he says.

Nevertheless, a one-gigabit-chip project would call for an unprecedented level of cooperation, industry leaders concede. Although the technical hurdles of building a one-gigabit memory chip are significant, Meindl says, "the sociological problems are even more formidable." —*Elizabeth Corcoran*

## Sound Bytes

*Electronic music gains a human touch*

Somewhere, off in the distance, a gentle swishing begins to build, the sound of sand tumbling over a snare drum. Suddenly a sharp tinkling breaks in: tiny mallets striking crystal wineglasses. The blows fall more frantically, the clinking pitches higher—

While listening to composer Tod Machover's piece "Bug-Mudra," it is not difficult to keep in mind that much of the music is being generated by a computer. But watch Machover conduct the work on stage, and it is the integration of human and electronic performers that is unforgettable.

With his right hand, Machover directs a group of musicians with the familiar swings of a conductor. Meanwhile his left hand, encased in an enormous cybernetic "data glove" covered with sensors, directs a Macintosh II computer that controls a section of synthesizers. "My goal is to build things that make music become more creative, not just prosthetic devices," declares Machover, who is a professor at the Media Laboratory at the Massachusetts Institute of Technology. He currently records with Bridge Records in New York City.

Much electronic music continues to be developed precisely where it started—in the laboratories of computer scientists. But those who have spent years creating computer music, including Machover, are determined to keep people in the act. To do so, they are exploiting complex algorithms and new hardware to enhance the music performances of humans.

*HYPERCELLO, played by Tod Machover, is a hybrid of sensors and instrument that elaborates on the music being played. Photo: Peter Menzel.*

"Many music appreciators don't have the ability to play music but love it," says Max V. Mathews, a professor at Stanford University's Center for Computer Research in Music and Acoustics. More than 20 years ago Mathews helped to spark the synthetic music revolution with algorithms he devised at Bell Laboratories. Now he hopes to put music back into the hands of the listeners with his "radio baton." In many ways, it is a simpler version of the data glove used by Machover.

The radio baton is packed with several simple transmitters that send signals to an array of receivers that track the motion of the baton in three dimensions and communicate the data to a computer-simulated orchestra. A gesture to the left may enhance a bass line; a motion downward may speed up the tempo. "Instead of passively listening, people will be able to buy scores and conduct their own interpretations," Mathews predicts.

Others are trying to make computers more suitable performers in ensembles. "Our focus is to make the computer into a live instrument," explains Miller Puckette, a researcher at the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) in Paris.

Puckette and his colleagues have been developing algorithms that enable a computer to track the notes being played by a nearby live instrument and calculate its place in the music score. "Finding the pitch is a terrible analysis problem," he says. IRCAM is also building algorithms that let the computer follow the leads from other instruments. After picking notes from an ongoing performance, other, established algorithms will generate new combinations—chords or arpeggios, for example—around them.

Several investigators are developing "listening assistants"—computer programs, including neural networks, that follow music as it is being played and learn ways of improvising from the underlying structure of the piece. These days, David Wessel, who directs the Center for New Music and Audio Technologies at the University of California at Berkeley, is often found playing jazz-inspired improvisations with two musicians and a computer-based listening assistant. "The assistant takes fragments of what they're playing, elaborates on them and reinjects them in an immediate way into the performance," Wessel says.

Similarly, the "hyperinstruments" being built by Machover and other researchers at M.I.T. interpret an ongoing performance in novel ways. Musicians play hyperinstruments much as they would acoustic ones. The electronic hybrids, however, are packed with sensors that pick up every inflection of the performance. The data are the basis for new synthesized sounds that enhance the performer's original expressive intent.

So that computers will be regarded as performance instruments in their own right, investigators are also looking for ways to "recapture the intimacy of sound" by incorporating the nuances of acoustic instruments, Wessel says. Push a bow across the string of a violin, for instance, and the note produced will be a rich collection of sounds: the initial attack followed by a crescendo that fades to a subtle buzz. Computer-driven synthesizers, in contrast, have typically produced more uniform sounds. "The disadvantage is fixed in the hardware," says Carla Scaletti, president of Symbolic Sound Corporation in Champaign, Ill.

Typically, conventional music systems store a collection of digitized sounds—say, all the notes played by a piano—or a specific algorithm that dictates how to combine signals. Call for middle C, and the computer sends the binary description of the note to a synthesizer. What Scaletti and others aim to do is replace that library of digital signals with digital-signal processors that can be easily reprogrammed by a user. In this way, a person could redesign C or build a completely new sound.

So far synthesizers based on digital-signal processors are still expensive and tricky for musicians to use, experts say. But a handful of companies are beginning to develop commercial products. Early this year Symbolic began selling Kyma, a hardware and software system based on as many as nine digital-signal processors. Kyma includes 80 initial "sound objects." With an icon-driven program, a user can create a palette of other sounds by performing various operations on the data. "This is set up for exploring—not for someone under intense time constraints," Scaletti cautions.

Advances in digital-signal processing will "open doors for us to really think about music in different ways," Machover adds. Musicians, in turn, must convince audiences that there is substance and depth behind the novel orchestrations. One opportunity will take place soon: in August cellist Yo-Yo Ma debuts Machover's latest work, a 25-minute piece in three movements for three hypercellos at Tanglewood in Lenox, Mass. The audience's reaction will, of course, be live. *—Elizabeth Corcoran*

## FORTRAN Forever

*Is it still the language of choice for science?*

If you wanted to program a computer for scientific or engineering problems during the 1960s, FORTRAN was the language you used. It was clunky and unforgiving, but it was all there was.

Then computer scientists came up with languages that were easier to use, more elegant and more powerful: C, BASIC, APL, LISP, PASCAL and others. But physicists and engineers kept right on using FORTRAN. They analyzed stresses on bridges and aircraft, modeled the inside of fusion reactors and predicted the evolution of supernovas, all in a language firmly tied to punch cards.

With the advent of multiprocessor computers that break problems into parts and solve each section simultaneously, language designers began whipping up a whole new set of computer dialects. For these parallel processors, they invented SETL, NETL, ACTORS, LINDA and more. But recently a group of software engineers completed a 13-year effort to standardize an older computer language they claim is almost perfectly suited for a wide range of parallel programming tasks.

What's it called? FORTRAN. Well, FORTRAN 90, also known as FORTRAN EXTENDED. (Previous versions were FORTRAN 66, which was standardized in 1966, and FORTRAN 77, which was actually approved in 1978.) The international version is done, and approval of a U.S. standard is expected by the end of June.

The renovated FORTRAN was really designed to make it easier to write scientific or engineering programs. But by a stroke of good fortune, the same features that simplify life for programmers also furnish clear guideposts for running the programs on machines incorporating hundreds or even thousands of processors, says Danny Hillis, the parallel computer designer and founder of Thinking Machines Corporation in Cambridge, Mass.

The most important change in FORTRAN 90 is new instructions that can manipulate arrays of data—for instance, the temperature and wind velocity at every point on a grid covering the Western Hemisphere—as a single unit. A compiler (the computer program that translates FORTRAN statements into binary code) can turn such statements into commands that carry out operations on each element of the array in parallel. In the past, programmers had to write instructions to handle each data item individually.

Geoffrey C. Fox, a computer scientist at Syracuse University, estimates that FORTRAN 90 will be nearly perfect for the parallel programming of about half of all the scientific and engineering computing done today. And many other programs can be recast to fit the kind of parallelism that FORTRAN 90 supports, he says.

The evolution hasn't been easy. Work on updating FORTRAN has been going on for nearly 14 years. Now although a number of upstart computer companies, including Thinking Machines, MasPar, Alliant and AMT, already have compilers that incorporate much of FORTRAN 90, older, more established firms have been dragging their feet, Hillis says.

Moreover, some programmers—the ones whose life FORTRAN 90 was intended to simplify—are not rushing to embrace the improved version. They complain that it bears little resemblance to earlier versions and has features that will be difficult to learn. Fortunately, users can take their time learning the new dialect. Old programs will still run under the new standard, says Jeanne Adams of the National Center for Atmospheric Research in Boulder, Colo., who chairs the FORTRAN 90 standards committee at the American National Standards Institute.

But if they take too long, FORTRAN may change again. Fox and others foresee extending FORTRAN 90 in a few years to increase the range of easy-to-write parallel programs and to put back features removed from early drafts of the language by conservative standard-setters. How many times can the venerable language be updated? Computer-language designer Guy L. Steele of Thinking Machines quotes an unknown scientist: "I don't know what the computer language of the year 2000 will look like, but I know that it will be called FORTRAN." *—Paul Wallich*

## 3-D Documents

*Holograms may add dimension to computer data*

Seeing a house that Frank Lloyd Wright designed but never built is a breeze with a graphics workstation. Just tell the program what to do, and you can change the lighting to see the house at dawn, in the glare of noon or as the shadows lengthen at sunset and the lights come on. If you want to view the structure from another angle, or step inside, the computer will accommodate you. The catch is, you can't take those changing three-dimensional images with you.

That bothered John R. Andrews, a member of the research staff at Xerox. How convenient it would be, he thought, to be able to slip animated three-dimensional documents into a briefcase—not just architectural drawings but everything from the workings of a turbine to ink flow from a printer.

Andrews and his colleagues at the Xerox Webster Research Center in upstate New York have found a solution. They have devised a system for converting computer graphics into animated holograms, three-dimensional images that change in time as the viewer's angle of vision changes. Now Andrews anticipates the day that office computer printers will produce large-format holograms for leisurely review. "These holograms will add a new dimension to the classical format we think of as the document," he declares.

To produce the holograms, Xerox encodes the light coming from a computer screen into the optical interference patterns that will form a hologram when illuminated from behind. Each frame is recorded on clear film as a narrow horizontal strip. A complete animation, such as that illuminating the Wright house from dawn till dusk, can contain as many as 70 strips.

The resulting stack of images resembles a venetian blind. A viewer could see each frame, but the effect would be like peering at the scene through a tiny crack in the blind. A very tiny crack, in fact, because the slits are kept to the average width of the eye's pupil, about three millimeters. So that the entire scene can easily be viewed, Xerox creates a second hologram of the first. The copying process renders the strips invisible by moving them out of the eye's focus—it is as if the blind were pulled up completely. The observer receives variable information such as time or space by moving up and down in front of the illuminated image.

The complexity of the process remains an obstacle to commercialization. To simplify production, Lambertus Hesselink, an applied physicist at Stanford University, suggests using a hybrid system for computing and recording digital holograms, such as the one Stanford has patented. Instead of an entire screen being calculated point by point, it is divided into sections, like a mosaic. For each patch, the computer calculates the physical characteristics of light that would bounce off an object if it existed outside the computer. "Then we implement what the light should look like with a laser beam and shine it on that patch," Hesselink says.

No one will speculate when computer printers will begin churning out three-dimensional animations, but Hesselink and others agree that it is certainly possible. The first step, they predict, will be less ambitious than Xerox's scheme—probably just single three-dimensional images. But even that promises to provide a new window into computer data. *—Deborah Erickson*

---

## A Matter of Taste

*Now holographers bring you food with a view*

Would you want a chocolate bar that flashes like a crystal chandelier? How about an Easter bunny whose eyes follow you around the room? Eric Begleiter says he does, and he bets a lot of children do, too. "We did focus groups and found that kids want candy that's a little spooky," he says.

Begleiter is president of Dimensional Foods Corporation of Boston, a company he founded in 1987 to promote his patented process to festoon foods with holograms. Begleiter is no stranger to the strange. Having won his spurs at the Center for Advanced Visual Studies at the Massachusetts Institute of Technology, he and a colleague once proposed staging a celestial light show out the back of a space shuttle decked with electron guns. Some of his M.I.T. associates have invested in his firm, he notes proudly, along with several straight and narrow candy-industry executives.

Holograms encode optical information in dizzying microscopic patterns. The most elaborate ones capture several sides of an object in a true three-dimensional image or animate them in sequences that do not require multiple frames. Simpler patterns suffice to create flashes of color (a selling point in the dye-shy health-food market).

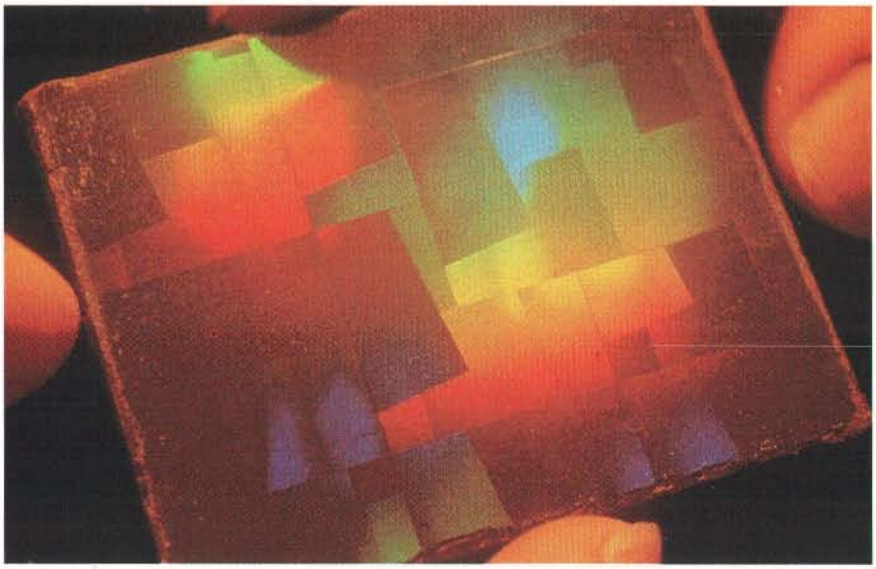
Abstract holographic patterns can be transferred directly onto the candy as it cools in etched plastic molds. Full-blown images, such as a rose that appears to rise from within a chocolate bonbon, can also be formed by rippling ridges, whose one- and two-nanometer size makes them too tiny to rasp a tongue.

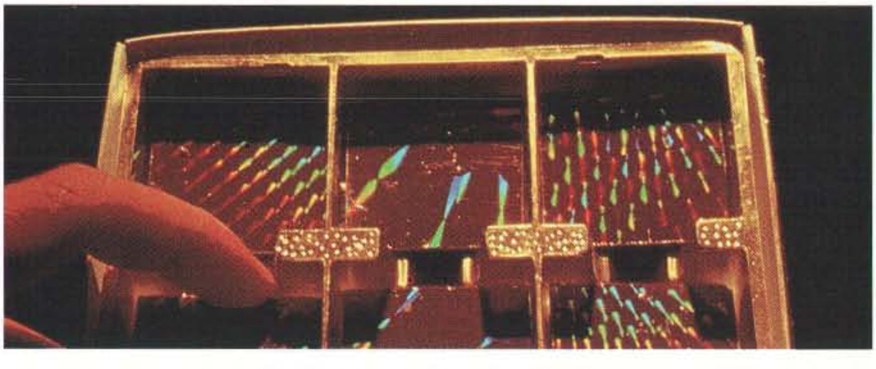
"It's surprising, but surface rigidity is not an impediment," Begleiter says of the candy on which he prints his holograms. "We only want the surface hard enough to retain the print to market, with good mouth feel." Any candy hard enough to hold a dull gloss will work. Chocolate retains the pattern for months if kept cool, he says, and shelf life can be extended by adding a protective sugar coating.

Begleiter's first market-ready product is mere frippery, a kind of pixie dust that can be added to corn flakes or sprinkled on cake. It consists of confetti made from crumbled sheets of colorless, tasteless, transparent starch that carry light-diffracting reflective holograms—the kind printed on credit cards as antiforgery devices. Begleiter says the sparkles go to market this fall but would not name the food companies that are licensing the process.

Dimensional Foods says it will offer a broad repertoire of three-dimensional chocolate patterns by the end of next year to retailers that make candies for such holidays as Easter, Halloween and Valentine's Day. Potential markets are presented by *chocolatiers* that serve hotels and other customers who crave distinction. Holograms also make edibles hard to tamper with, a feature that Begleiter says has interested pharmaceutical companies that make pressed-powder tablets (caplets).

Transmission holograms that create three-dimensional images inside transparent candies are the next step, Begleiter says. One experimental lollipop shows animated cartoons when it is twirled; another sucker is being designed to melt its way through a series of concentric images. The structures of the eye, perhaps? *—Philip E. Ross*





*KALEIDOSCOPIC CHOCOLATE* **(top)** ***reflects light from the tiny ridges of a holographic pattern. The ridges are transferred as the candy cools in molds*** **(bottom).**

# THE ANALYTICAL ECONOMIST

## Games That Networks Play

There was a time in the 1970s when it seemed that economists believed they could invoke game theory to predict decisions in any transaction. The idea was simple: describe the responses of two parties to a proposed deal by a matrix in which the rows correspond to choices by one player and the columns to choices by the other. For every combination of choices, there would be a payoff. Problems involving more than two people required more complicated matrices. If players worked to maximize their own expected payoffs, the combined choices would correspond to the Nash equilibrium, a point beyond which no player could improve his or her situation by unilateral action.

But for economists, game theory has largely proved to be more frustrating than helpful. "The hope that one could solve games and say economically useful things hasn't worked out," asserts W. Brian Arthur of Stanford University. Simple games such as Prisoner's Dilemma have straightforward solutions. But realistic games involving many players, each with a wide range of options, lead to multiple Nash equilibria and no way of predicting which will come to pass. Although the terminology that evolved from game theory has become widespread, Arthur maintains that it is "hard to find even one instance in economics" where simply setting up payoffs and solving the resulting game has yielded a useful analysis.

But the work of the economists on game theory is turning out to be applicable in a far different discipline: computer science. Researchers have been working for the better part of a decade to apply economic concepts to the problem of allocating such scarce resources as computer time, network bandwidth and access to file servers. Large networks containing a myriad of linked machines, all with different capabilities, defy easy analysis. In a classic "tragedy of the commons," many existing computer networks charge users only a flat fee for access, thereby inviting overuse.

In some experimental systems, computing jobs are allocated a certain amount of "credit," and idle central processing units offer their cycles to the highest "bidder." In others, networks charge different rates for different classes of service: digitized voice conversation, for example, might require a high-priority rate for immediate transmission, whereas electronic mail would cost less and accept delivery times measured in minutes instead of milliseconds.

Underlying these experiments is the assumption that the hidden hand of the market will inevitably allocate resources among greedy programs in a way that most nearly satisfies all of them—what economists call a Pareto optimal situation. But sometimes these "bit markets" turn out to be woefully inefficient. Greed is not enough, according to computer scientist Scott Shenker of the Xerox Palo Alto Research Center (PARC).

Shenker stumbled onto game theory while studying allocation problems. He used it to analyze the simplified problem of a computer-network gateway, which receives data packets from multiple users and transfers them to another net, queuing the packets up if they arrive too fast. Users can choose different amounts of transmission service; the "price" they pay for sending more packets is an increased delay.

Contrary to the hidden-hand assumption, Shenker found that a conventional gateway, which transmits packets on a first-come, first-served basis, could not guarantee anything like Pareto optimality. Users who transmit large amounts of data and are willing to put up with long delays slow the system down for everyone.

> *How will the hidden hand of the market allocate network resources among greedy computer programs?*

Game-theory analysis showed that another gateway algorithm, called Fair Share, does better. Fair Share transmits at least some of every user's data at high priority and additional amounts at succeedingly lower priority. This allocation is "envy free"—no user would want to switch places with any other—but it is not necessarily optimal. Indeed, Shenker is convinced that no gateway algorithm in his simplified market can guarantee an optimal solution. Some computer scientists see that as just one more proof that *radix malorum est cupiditas*—that greed is at the root of even computer bugs.

Other applications of game theory show that cooperative computer programs can allocate resources optimally. So why not make software that eschews greed? Shenker contends that building cooperative systems is impossible in practice. Agreeing on exactly how programs should cooperate means knowing in advance exactly what behavior will satisfy their users, he says. Furthermore, any universally standardized algorithm would become obsolescent and so prevent the efficiency it was trying to create.

Other researchers have also suggested that universally standardized systems are not feasible. Bernardo A. Huberman, a pioneer in applying principles of economics and ecology to computer systems, also at PARC, has shown that large networks will operate predictably only if they contain a heterogeneous population of machines—much as species diversity makes for resilient ecosystems. A network consisting of homogeneous systems, no matter how they try to cooperate, will eventually grow chaotic and break down, he says.

Economists who still support game theory seem pleased that it has taken on a career in computing. They also aver that there may be hope for it in economics as well. David M. Kreps of Stanford notes, for example, that some of the noncooperative game theory used by Shenker was first developed in studies of how firms vie for markets. Kreps also believes game theory can help economists think about such thorny problems as the economic effects of credibility—say, how Federal Reserve policy might depend on whether people believe Alan Greenspan when he insists that he won't lower interest rates any further.

In the meantime, economists of all persuasions will have to put up with flat-rate, congested computer networks just like everyone else.

*—Paul Wallich and Elizabeth Corcoran*

# Insectoids Invade a Field of Robots

Will the next major advance in robotics spring forth from inexpensive machines that crawl, think and act like bugs? Researchers in the "insect lab" at the Massachusetts Institute of Technology hope so. They have spawned a swarm of small robots that behave like your average arthropod. These insectoids, as I call them, are based on new principles of robot design and threaten a paradigm shift in the field of robotics.

Until recently, engineers bent on designing a robotic "brain" have taken a determinedly analytic approach. In this traditional view, they first decide what the robot will be able to sense; they then consider how it will analyze sensory inputs and finally how it will plan and take action. Each step is fraught with complexities that are likely to bog down intricate projects.

Abandoning the traditional approach, Rodney Brooks, director of the insect lab, has adopted a design philosophy that he calls subsumption architecture. To apply this philosophy, he starts by designing a network of processors and hardware that can produce a simple behavior. No behavior is added to the system until the behavior it subsumes is up and running (or walking, as the case may be).

For instance, to design an artificial creature that wanders and avoids obstacles, Brooks would first assemble a creature that moved randomly and then add the detectors and processors that would sense objects and instruct the creature to change direction. In subsumption architecture, complex behaviors evolve from a variety of simple features.

To test the practicality of subsumption architecture, Brooks and a team of his graduate students began building and designing many insectoids, from Allen, a primitive robot on wheels, to Squirt, a delicate bug no bigger than a grasshopper. But no creature illustrates subsumption architecture as well as Genghis, a foot-long assembly of motors, struts, gears and microchips.

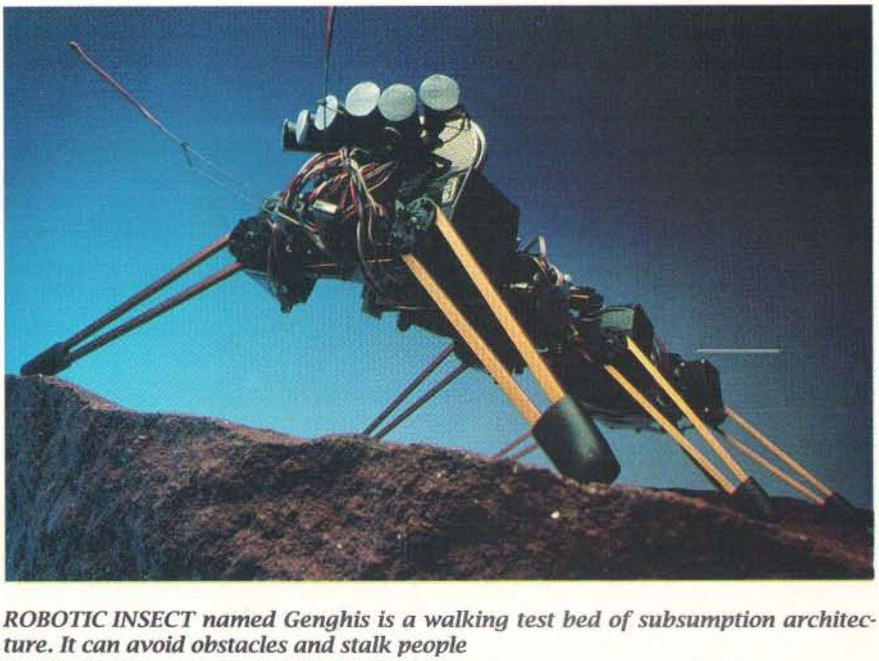
Genghis, who was created in part by graduate student Colin Angle, has six stiltlike legs, two whiskers and six infrared "eyes" transplanted from burglar alarms. Each leg is operated by a pair of motors [*see illustration on next page*]. An Alpha Motor moves the leg forward or back; a Beta Motor swings the leg up and away from the body or moves it down and toward the body. Between Genghis's legs are microchips that serve as the insectoid's nerve center. The microchips contain numerous augmented finite-state machines (AFSMs). Each AFSM stores numerical information for controlling various aspects of Genghis's behavior, such as the movement of legs. The information, or state, of an AFSM may change from time to time, depending on the input it receives from other modules. The state will also determine how it reacts to the inputs.

A robot must walk before it can run. In fact, to do anything worthwhile, Genghis must first stand up. Two numbers sent to the AFSMs control the Alpha and Beta Motors on each leg. One AFSM, called Alphapos, controls the Alpha Motor, and the other AFSM, Betapos, governs the Beta Motor. Each number represents vertical and lateral leg positions when Genghis is standing. As soon as Genghis is powered up in this simplest of behavioral settings, the motors all run until the leg positions (monitored by sensors) match the numbers stored in Alphapos and Betapos. The simple act constitutes what might be called the zero level of Genghis's architecture.

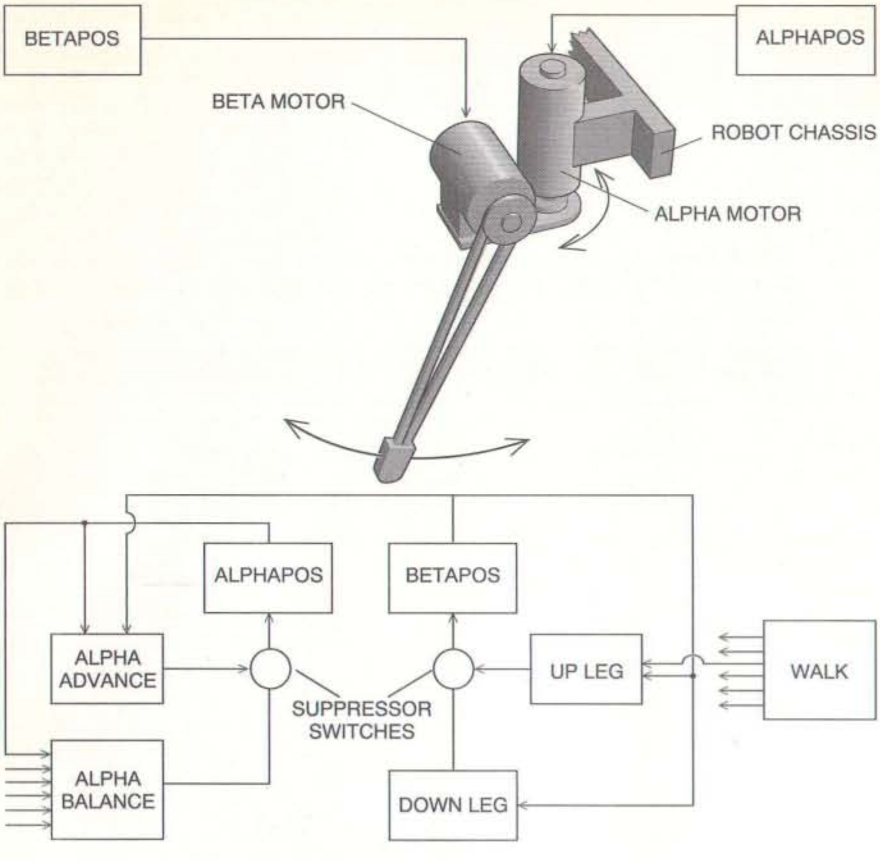
The next level of behavior, simple walking, is a feat robotics researchers have traditionally judged to be technically difficult. Genghis's basic walking network, in its simplest form, consists of two master AFSMs and 30 auxiliary ones, five per leg. Because the circuits for each of the insectoid's six legs are essentially the same, I will describe what happens to one leg and the five AFSMs that control it [*see illustration on next page*].

The key to basic walking is the global controller called Alpha Balance. This AFSM receives continual reports, in the form of numbers, on the positions of all six legs. A positive number indicates that a leg is pointing forward; a negative number, that it is pointing backward. Not surprisingly, legs that point straight out from the body are represented by zero. Alpha Balance adds these numbers together, the sum being a kind of average. If the sum is positive, it means that on average the legs are pointing forward. If the sum is negative, the average leg projects rearward.

The whole trick to walking revolves around the fact that if five of the legs touch the ground and a sixth is raised, then the insectoid may glide forward by a small amount merely by swinging all its ground legs slightly to the



***ROBOTIC INSECT named Genghis is a walking test bed of subsumption architecture. It can avoid obstacles and stalk people***



***BASIC CIRCUITRY that allows Genghis to stand*** **(top)** ***and walk*** **(bottom)**

rear. If the insectoid then swings the upraised leg forward and places it gingerly back on the ground, it is one small step for an insectoid but one giant leap for robotics.

When Ghengis swings a leg to the front, Alpha Balance generates a sum that is positive and then sends a negative signal to all legs that are currently down. Their motors whine briefly, the insectoid moves forward a bit and the signal is rebalanced. That is all the Alpha-Balance Module cares about.

The manner in which the various modules interact to create the act of walking amounts to an electronic ballet among the modules of the six leg networks. The action begins for a particular leg when its Up-Leg Module is activated. The activation sets off a chain of coordinated events among the modules; the Up-Leg Module then signals the Betapos Module, sending it a number that reflects an upraised leg position. The Betapos Module, which controls the Beta Motor, normally receives a positive number (that keeps the leg firmly planted) from another module called Down Leg. The new, negative signal from the Up-Leg Module suppresses the positive signal from Down Leg. Consequently, the Beta Motor raises the leg to a point where its reported position matches the new signal.

This event triggers a completion state in the Betapos Module, and it signals this state to three other modules: Alpha Advance, Up Leg and Down Leg. The Alpha-Advance Module, which controls the back-and-forth motion of the leg, sends a strongly positive signal to the Alphapos Module. The Alpha Motor whirs gently, and the leg waves forward, almost as if it were probing the air. When the Up-Leg Module receives the completion signal, its action is suppressed. When the Down-Leg Module gets the completion signal, it is activated, and the Beta Motor powers the leg down to terra firma.

A master module, called Walk, controls the entire movement by sending a sequence of signals to the six Up-Leg Modules. But what sequence should it use? The triggering pattern most commonly used by insects is called the alternating tripod gait. If the legs are labeled *R* for right and *L* for left, as well as numbered *1, 2, 3* from front to back, the alternating tripods are the sets *R1, L2, R3* and *L1, R2, L3.* In normal situations, an insect like a cockroach will lift the first set, *R1, L2, R3,* leaving the other set on the ground. This triangular stance gives stability to the cockroach as the first set of three legs swings forward to new positions. Then the other set can be raised and swung forward in the same way while the first set provides stability.

The Walk Module may send out a sequential version of this set of signals to the six Up-Leg Modules. Or it could send out the pattern sometimes used by stick insects: *R3, L1, R2, L3, R1, L2.* There are numerous possibilities for stable gait patterns.

Perhaps readers can figure out the gait of a millipede machine. If there are 1,000 legs on each side of an insectoid, devise a gait that will carry the creature forward without any leg getting dragged along by the body.

Using the primitive network just described, Genghis can walk but not very smoothly and not in the manner that Brooks describes as "robust." For one thing, Genghis wobbles excessively and cannot clear obstacles of even moderate height. The addition of a few more kinds of AFSM provides a new level of subsumption architecture and a new degree of behavioral competence.

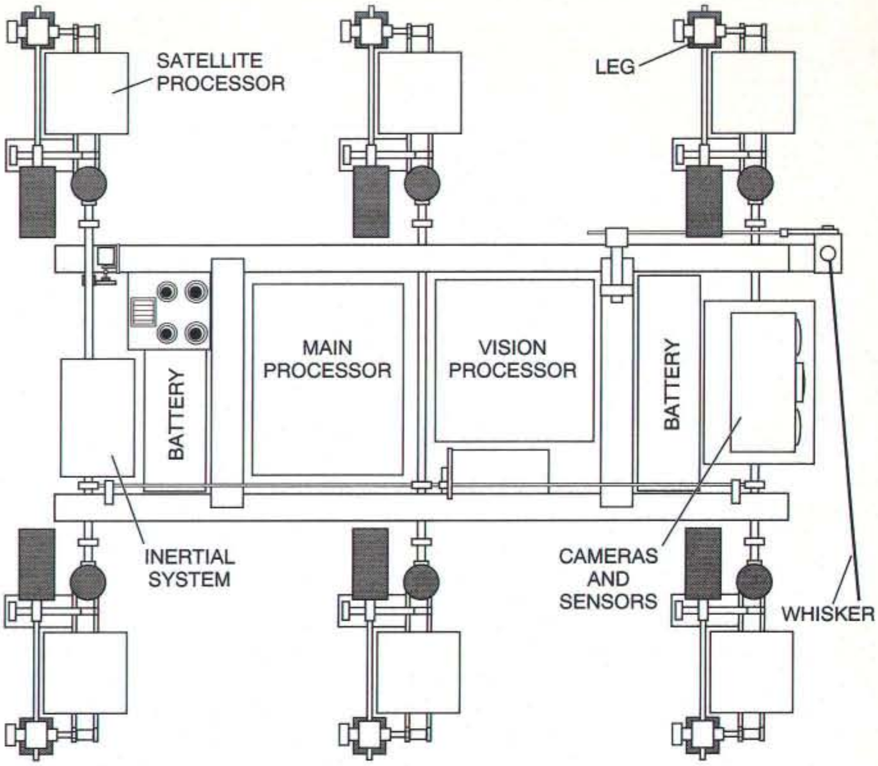
A Beta-Force Module monitors the high strain that develops in a Beta Motor when its leg has been set down in a position that supports too much of the creature's total weight. Genghis may have stepped on a five-centimeter rock, for example. The Beta-Balance Module for that leg senses the unusually high force and sends a zero message that suppresses the leg-down message and makes the offending leg "compliant." The leg, in other words, gives way a bit, and Genghis compensates for the high terrain under one of its legs.

But on sloping terrain, the downhill end of Genghis will take more force than the upper end, and the legs will become compliant, increasing the pitch even more. Correcting this problem requires two Pitch Modules to monitor the outputs from a pitch-measuring device. The Pitch Modules send messages to inhibit whichever Beta-Balance Modules have become too compliant.

When Genghis encounters an obstacle while swinging one of its legs forward, a sensor on the motor picks up the additional strain and sends a message to an Alpha-Force Module. This AFSM then sends a signal to the Up-Leg Module, which then results in a higher leg lift.

Among the many sensors used by Genghis are two whiskers and six infrared sensors. The whiskers send their reports to a feeler module. If a whisker senses an obstacle, the feeler module resets the Up-Leg Module for one of the two front legs.

The infrared sensors introduce the next major level of subsumption. The



***ANATOMY of an insectoid named Attila is shown above and at the left***

sensors work in conjunction with a Prowl Module, which gives Genghis a somewhat sinister mode of behavior. In this mode, Genghis rests quietly until it detects infrared radiation from, for example, a nearby human ankle. When that happens, Genghis activates its Walk Module. The creature then begins to creep forward like some demented insect toward the hapless human. Of course, there is plenty of time to get out of its way, but if a Steer Module is added as well, Genghis can be relentless.

A year or so ago, when a curious visitor saw the insectoid for the first time, he asked, "Is it a bug?"

"No," Brooks said, repeating an old programming joke. "It's a feature." For a while, Brooks insisted on calling the insectoid "Feature." But later a graduate student suggested "Genghis," which seemed more appropriate for a creature whose instincts were to stalk and conquer.

Recently some new circuits were added to Genghis to test whether self-organizing behavior might emerge in the absence of a central control module like Walk. The results were impressive. The microchip ganglion associated with each leg was given the option of running its own experiments with a set of basic behaviors like lifting or lowering a leg or swinging it forward or back. Each experiment consisted of recording what the neighboring legs were doing, then trying one of the basic behaviors and checking whether the body fell down or not. Fascinating to watch, according to Brooks, the experimenting insectoid might sit for a while, legs waving in the air, next thrash for a bit, then begin to move forward with tentative steps. Within a minute and a half, the network always "learned" the alternating tripod gait!

The notion of autonomy dominates the subsumption approach to robotic architecture in the M.I.T. insect lab. Can a robot, no matter how small, be given a behavior that will enable it to survive in the real world for extended periods? The insectoid called Squirt will fit inside a one-inch cube. Too small for legs in the current state of insectoid technology, it features wheels, a single motor, a microprocessor, two lithium batteries and three sensors. It uses two microphones to listen for sounds and a single light sensor to gauge the amount of light available.

Squirt will survive, provided it does not get stepped on. For this reason, it has been programmed with several layers of behavior that were transferred electronically from a computer to its single microchip. Squirt hides in the dark while listening for sounds. If it hears nothing for a few minutes, it ventures out in the general direction of the most recently heard sound. After wandering for a while, it engages in a spiral search to find a new hiding place.

In this respect, Squirt resembles the vehicles imagined by the German scientist Valentino Braitenberg [see "*Braitenberg memoirs: vehicles for probing behavior roam a dark plain marked by lights*," "Computer Recreations," by A. K. Dewdney; SCIENTIFIC AMERICAN, March 1987]. The purpose of Braitenberg vehicles, among other things, was to illustrate the thesis that very complex behaviors could result from very simple control systems. The thesis has inspired more than one robotics enthusiast to build a behavioral vehicle. But it was Brooks and company, also intrigued in part by Braitenberg's vehicles, who succeeded first. Some might be willing to ascribe emotions such as fear or longing to the behavior of the neurally controlled vehicles. Is Squirt afraid of people? It certainly acts like it.

Graduate student Anita Flynn, who was part of the team that built Squirt, sees the future of robotics blossoming in even smaller insectoids she calls gnats. These creatures would be the size of real insects, not to mention gnats themselves. Their body parts would be fabricated by the same techniques currently used to etch microcircuits on silicon surfaces. The biggest bottleneck is the ultratiny motors that gnats will require. The field of microengineering has already produced gears that would scatter at a sneeze.

Will the insectoids rule some day? Brooks is cautious about making claims about the future of subsumption architecture, but he plans to push the idea to its limits. Will we find that as we add ever more complex layers of behavior that the subsumption approach will continue to work? Or will we encounter a barrier that forces us to resort to something like traditional techniques? I am sure these questions will bug him for the foreseeable future.

FURTHER READING

GEARING DOWN. Ivan Amato in *Science News*, Vol. 139, No. 2, pages 26–27; January 12, 1991.

BEAM ROBOTICS. Michael C. Smit and Mark W. Tilden in *Algorithm*, Vol. 2, No. 2, pages 15–19; March 1991.

# BOOK REVIEWS

by Philip Morrison

## The Robust Will to Believe

**FANTASTIC ARCHAEOLOGY: THE WILD SIDE OF NORTH AMERICAN PREHISTORY,** by Stephen Williams. University of Pennsylvania Press, 1991 ($28.95; paperbound, $14.95).

Sweetness and good humor grace this well-illustrated work by a senior Harvard archaeologist/anthropologist, teacher and museum curator. Without those traits in abundance, this questioning journey along the wild shores of wishfulness might leave a reader little but the sourness of disillusion. Williams, however, starts with a generous premise: "Without fantasy, science would have nothing to test." Curiosity and the imagination it kindles come first, but it is stringent testing and veracious reporting that build a science.

A dozen chapters of narrative carry us delightedly among the outlandish instances of two fantastic centuries. The oldest example in this genre of American Humbug comes from a real monument, the Grave Creek Mound, a "lofty and venerable" tumulus not far from the banks of the Ohio River, 12 miles south of Wheeling. The landowners excavated their 70-foot earthen mound in 1838, to disclose a number of burials amid grave goods, the copious shell beads and cut mica of what we call the Adena culture, which we date now to about 400 B.C. The burials were a royal personage and his retainers, said the scholars of the time, citing Scythian burials from Herodotus and more recent finds in Eastern Europe. We cannot add much more in our day, in spite of many other Adena finds.

A few years after that mound was opened, a celebrated traveler called attention to a wonder ignored among the broken stone tools and minor relics housed nearby. It is not clear where or when it had first appeared, probably in a wheelbarrow load of dirt. But it was unmistakably an inscribed stone, bearing some 25 characters in three lines, and a minor drawing. More than one X can be made out; surely this is alphabetic. It remains unique, for those people left us no other inscriptions.

We were told what it says, as construed by more than one scholar over the ensuing decades. One pundit made out: "The Chief of Emigration...has fixed these statutes forever." Another read, "The Grave of one who was assassinated here. May God to avenge him strike his murderer." There is a version, doubly translated by way of the French: "What thou sayest, thou dost impose it,... in thy impetuous clan." Perhaps; no linguistic process is cited, nor would one be much open to lay criticism.

But this world is not made of scholars' words on paper. We have the Grave Creek Stone itself. It is no finished tablet of hard rock engraved by the craftsman's skill, only one little sandstone skipping-pebble less than two inches long. The marks could have been made in a few minutes with nail or knife point by anyone at hand in the season that high tumulus was opened.

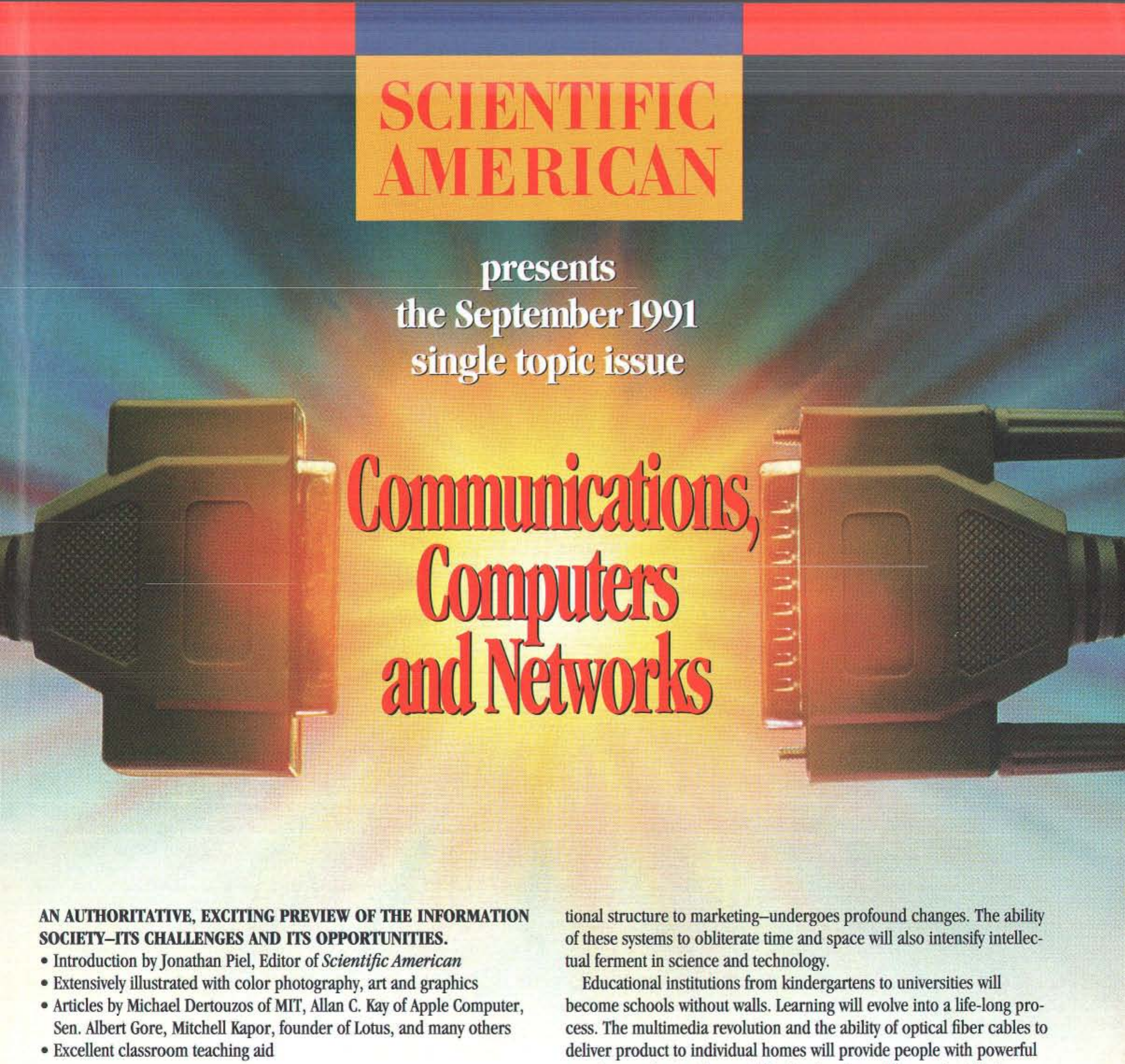
Nor can this wild romance be dismissed wholly as the prejudiced groping of a naive century, well past. A fourth rendering was published in 1976, over the imprint of a well-known New York publisher, by a tireless modern epigrapher, once a Harvard biologist. Dr. Barry Fell finds the script to be an alphabet used in Spain in the first millennium B.C., the language Phoenician. Its message: "The mound raised-on-high for Tasach / This tile / [His] Queen caused-to-be-made." (Plainly the Queen's mound-raising budgeters had skimped on this particular line item.)

*FRAUDULENT "Michigan Relic" shows Noah's Ark and inscriptions. The supposedly millennias-old artifact was probably made around 1910 by James O. Scotford, whose mark is at the upper right.*

Take an example from the cultivated 1890s. A Harvard chemistry professor grown wealthy out of the success of a patent baking powder moved to a fine Cambridge home near the Longfellow house. The man was inspired to seek the Vinland of the sagas by a visiting Norwegian celebrity. Stroll today on a busy riverbank street a few blocks away, and you will find the monument he caused-to-be-made in stone and bronze a century ago to mark the nearby site of the very house of Leif the Lucky. Luckily indeed, long-sought Vinland turned out to be right in that same livable neighborhood. A few colonial artifacts were found by digging under the walls of one 18th-century structure. This enthusiast defended his finds in half a dozen learned books, even to the inference of a Norse city of 10,000 Vikings built a few miles upriver, long vanished without a trace.

Since 1960 we have known that the Vikings were really ashore here in North America; a cluster of a few ruined sod houses with their homely artifacts—no fine masonry towers or advertisements in stone—have been found just about where you might expect, in northern Newfoundland. They disclose a concordant carbon 14 date and a strong similarity to well-known archaeological finds in Greenland. It was never the possibility of a transient Viking presence the skeptics had stoutly denied, but the faulted arguments offered in support of what the sagas hinted; today no one rejects the complex at L'Anse aux Meadows. It is process that validates, and not outcome. Nothing is too wonderful to be true—*if* it is true.

What a bundle of wishes, folly and fraud is here unpacked! Great cities swallowed beneath the sea in Atlantic and Pacific both, sacred revelations deciphered from vanished tablets or even from visions, and polyglot emigrants crowding to the New World from all quarters of the globe. To Ohio they



**AN AUTHORITATIVE, EXCITING PREVIEW OF THE INFORMATION SOCIETY–ITS CHALLENGES AND ITS OPPORTUNITIES.**

- Introduction by Jonathan Piel, Editor of *Scientific American*
- Extensively illustrated with color photography, art and graphics
- Articles by Michael Dertouzos of MIT, Allan C. Kay of Apple Computer, Sen. Albert Gore, Mitchell Kapor, founder of Lotus, and many others
- Excellent classroom teaching aid

As computer networks spread, we will find ourselves facing major opportunities and challenges as every facet of business–from organizational structure to marketing–undergoes profound changes. The ability of these systems to obliterate time and space will also intensify intellectual ferment in science and technology.

Educational institutions from kindergartens to universities will become schools without walls. Learning will evolve into a life-long process. The multimedia revolution and the ability of optical fiber cables to deliver product to individual homes will provide people with powerful new forms of entertainment. The legal system will grapple with the challenge of reformulating public policy, property laws and civil liberties.

☑ **Yes,** I would like to receive the single topic issue: *Communications, Computers and Networks.*

Name ______________________

Company ______________________

Address ______________________ Apt

City ______________ State ________ Zip

Subscribers will receive this single topic September issue as part of their regular subscription.

ON SALE
AUGUST 27th
AT YOUR
NEWSSTAND
FOR ONLY $3.95

Please send: ________ copies at $3.95 ($4.95 outside the U.S.) $________

Add $1.00 per copy for postage and handling $________

TOTAL ENCLOSED: $________

Postage and handling paid by *Scientific American* on orders of ten or more copies. Please make check or money order payable to: *Scientific American.* Send orders to:

Scientific American
Dept CCN
415 Madison Avenue
New York, NY 10017, USA

CCNSA7

**SCIENTIFIC AMERICAN**
**In Other Languages**

**LE SCIENZE**
L. 5,000/copy L. 50,000/year L. 65,000/(abroad)
Editorial, subscription correspondence:
Le Scienze S.p.A., Piazza della Repubblica, 8
20121 Milano, Italy
Advertising correspondence:
Mondadori Pubblicità S.p.A., Via Rivoltana 8
20090 Segrate, Italy

**サイエンス**
Y 980/copy Y 10,753/year Y 13,393/(abroad)
Editorial, subscription, advertising correspondence:
Nikkei Science, Inc.
No. 9-5, 1-Chome, Otemachi
Chiyoda-ku, Tokyo, Japan

INVESTIGACION Y
**CIENCIA**
500 Ptas/copy 5500 Ptas/year 6200 Ptas (abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica, S.A.,
Viladomat, 291 6ª 1ª
08029 Barcelona, Spain

POUR LA
**SCIENCE**
29 FF/copy 296FF/year 350FF/year (abroad)
Editorial, subscription, advertising correspondence:
Pour la Science S.A.R.L.,
8, rue Férou,
75006 Paris, France

**Spektrum**
DER WISSENSCHAFT
10.50DM/copy 105.60DM/year 115.20DM/(abroad)
Editorial, subscription correspondence:
Spektrum der Wissenschaft
Verlagsgesellschaft mbH
Moenchhofstrasse 15
D-6900 Heidelberg
Federal Republic of Germany
Advertising correspondence:
Gesellschaft für Wirtschaftspublizistik GWP mbH
Kasernenstrasse 67
D-4000 Duesseldorf,
Federal Republic of Germany

**科学**
3.80RMB/copy 45.60RMB/year $48/(abroad)
Editorial subscription correspondence:
ISTIC-Chongqing Branch, P.O. Box 2104,
Chongqing, People's Republic of China

**В МИРЕ НАУКИ**
2R/copy 24R/year $78.50/(abroad)
Editorial correspondence:
MIR Publishers
2, Pervy Rizhsky Pereulok
129820 Moscow U.S.S.R.
Subscription correspondence:
Victor Kamkin, Inc.
12224 Parklawn Drive
Rockville, MD 20852, USA

**TUDOMÁNY**
98Ft/copy 1,176Ft/year 2,100Ft/(abroad)
Editorial correspondence:
TUDOMÁNY
H-1536 Budapest, Pf 338
Hungary
Subscription correspondence:
"KULTURA"
H-3891 Budapest, Pf. 149
Hungary

**العلوم**
1KD/copy 10KD/year $40/(abroad)
Editorial, subscription, advertising correspondence:
MAJALLAT AL-OLOOM
P.O. BOX 20856 Safat,
13069 – Kuwait

Advertising correspondence all editions:
SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue
New York, NY 10017
Telephone: (212) 754-0550 Telex: 236115

brought small stone tablets in modern Hebrew and stone pipes carved into the form of elephants; to Arizona, Roman swords cast out of lead and one entire Viking craft beached in the Mojave desert! (O Californians, by the worst of misfortunes it was buried there under a seismic landslip before anyone could revisit it.)

Sometimes baser motives obtrude. The Moundbuilders could not have been the Indians then being dispossessed, but another worthier, literate stock. The off-and-on finds of Norse relics peaked in a loyal Scandinavian reaction to the heated celebration of the 400th anniversary of Columbus's first voyage in 1892. A Viking longship sailed proudly into Chicago to the World's Columbian Exposition, preempting the arrival of three caravels in replica. Uncertain rune stones and suburban Vinland followed soon after.

Yet low origins cannot undermine real worth, as the Ninja Turtles are taught. A tendentious purpose does not of itself cancel evidence, although it makes airtight the case for strict testing. Appeals to consistency, context and chance all fall before the robust will to believe; this engaging and compassionate volume of spectacular examples offers ample proof. Surely the proper stance must not sanctify Authority, either; complacency is no more acceptable in science than is credulity. We will surely see changes even in what all experts now accept, but changes controlled by well-sifted new evidence.

An epilogue wakes us from these dreams; it is the author's eloquent summary of the actual prehistory of the Native Americans as we now see it. The widest cultural diversity is exhibited from the Arctic to Cape Horn, a "real fantasy" both of glory and of tragedy, still incomplete, still full of puzzles. The final annotated list of readings is a first-rate guide to deeper study; the principal "rogue professors" and other speculators in fantasy are fairly represented, from P. T. Barnum to yesterday's keen dowser.

## Necessary White Lies

**HOW TO LIE WITH MAPS**, by Mark Monmonier. University of Chicago Press, 1991 ($27.50; paperbound, $12.95).

"Not only is it easy to lie with maps, it's essential." The map paradox is intrinsic: "To present a useful and truthful picture, an accurate map must tell white lies." In a brief, informal survey, this Syracuse University geographer instructs and entertains, occasionally with a cynic's tongue in cheek. He explains those necessary white lies of cartography and goes beyond to both contrived and real examples of grayer lies, even to black ones, clarifying the real power of the cartographic abstraction, both its "enormous benefits" and its genuine costs.

The simplest elements of the map, its scale, projection and symbolization, are each a source of distortion. Projections are particularly well treated here without equations by invoking the use of a two-stage process. First, the round earth is thought of as a sphere of smaller size. Then that sphere surface is mapped, usually by some systematic procedure onto a flattenable plane, cone or cylinder.

The Mercator projection is the right map for one who steers by compass. On it every straight line marks a course that can be followed by simply holding to a fixed angular bearing measured from your north-south line. On another simple projection, called central or gnomonic, a straight line marks instead the shortest path between two points, the great circle. En route the compass bearing varies continuously. But these enforced compromises are marked by their complementary quality. A practical navigator can locate a few way points on a gnomonic chart, enter them on the Mercator map and strike simple bearings from each way point in turn. Those line segments approximate the great circle. Each map is in a way a theory that favors certain approximations. Procedures like selection, simplification, smoothing, displacements to make room, and out-of-scale notation for bridges, streams and roads so narrow that they would become invisible at true scale enter inescapably.

Mercator's easily laid out world projection was the classroom wall map standard for a long time and is still the backdrop to some video news. It sends the poles flying to infinity and enlarges areas nearby. Certainly a Mercator world map flattered the British Empire with a giant red Canada: some maps even repeated Australia and New Zealand on both edges of a Greenwich-centered world. Twenty years ago the historian Arno Peters presented his "new" equal-area world map, its long tapering southward continental finger forms in fact based on one drawn in 1855. "Dr. Peters knew how to work the crowd" and proclaimed worldwide that his map ended the stagnation of First World cartography and brought new fairness to the South. Many organizations concerned with development adopted and circulated the novelty, enraging earnest cartographers by their simplemindedness and ahistorical claims. They paid a high opportunity cost, for they might have adopted an eloquent if tailored projection on which the areas of nations are scaled to the numbers of their people. A small one shown here is striking, with Java a huge island, continental India and China dominant. (No date is given for the census, but Mexico, Nigeria and Egypt are all growing apace.)

Disinformation on maps, not merely misleading design, is far from rare. Soviet maps moved towns around between editions, presumably to fool targeters abroad. British maps bear wonderful detail, yet omit or mislabel such security entries as radio stations or government oil depots. The Greeks, more candid, issue maps with large blank areas. In one of the author's ironic lists, rules for "polishing the cartographic image," we meet "the architect's cartographic friend, the tree stamp.... symbolic trees can convert a mundane proposal into a...neighborhood asset." A few published pranks by unknown map drafters are shown.

The most frequent theme is the problem of maps for which some aggregation of detail is necessary. It is sobering to see examples that demonstrate how misleading the same data can be when differently mapped. John Snow dotted a street map with the locations of all recent cases of cholera in one district of Victorian London. They clustered neatly around the Broad Street pump, a public water source; once its handle was removed, new cases nearby became rare. Three gray-scale maps produced by different areal aggregations of Dr. Snow's dots all obscure the inference.

Professor Monmonier himself knows how to gain our attention; it is not in fact the lies in maps but their truth, if always approximate and incomplete, that he wants us to admire and to use, even to draw for ourselves on the facile screen. His is an artful and a funny book, which like any good map packs plenty in little space.

## Counterforce

**INVENTING ACCURACY: A HISTORICAL SOCIOLOGY OF NUCLEAR MISSILE GUIDANCE**, by Donald MacKenzie. MIT Press, 1990 ($29.95).

The black boxes able to steer any craft or missile sit at the intricate apex of contemporary technology, although conceptually they arise in simple Newtonian mechanics. Two step-by-step summations over time of the changing vector acceleration you

Discovery.
Exploration.
Cooperation.
These are the
hallmarks of this
planet's increasingly
international advance
into space—and for
9 days in 1992, the
World Space Congress
will mark the International
Space Year with the most
significant gathering
of space scientists
and engineers
in history.

**THE WORLD SPACE CONGRESS**

WASHINGTON, DC • 28 AUGUST–5 SEPTEMBER 1992

**A CALL FOR PAPERS**

**The Historic World Space Congress** issues an official *Call for Scientific and Technical Papers* to be considered for presentation at the first-ever joint meeting of the 43rd Congress of the International Astronautical Federation (IAF) and the 29th Plenary Meeting of the Committee on Space Research (COSPAR). The World Space Congress is hosted and organized on behalf of the United States by the American Institute of Aeronautics and Astronautics (AIAA), and held under the auspices of the National Academy of Sciences (NAS) and the National Aeronautics and Space Administration (NASA).

To obtain the World Space Congress *Call for Papers* booklet, program overview, and general information, including registration forms, hotel, and travel details, write:

WORLD SPACE CONGRESS 1992

World Space Congress
c/o American Institute of Aeronautics and Astronautics
The Aerospace Center, 370 L'Enfant Promenade, SW
Washington, DC 20025-2518
202/646-7451 (World Space Congress Hotline)
202/646-7508 FAX

Authors...
LOOKING FOR A PUBLISHER?
Learn how to have your book published.

You are invited to send for a free illustrated guidebook which explains how your book can be published, promoted and marketed. Whether your subject is fiction, non-fiction or poetry, scientific, scholarly, specialized, (even controversial) this handsome 32-page brochure will show you how to arrange for prompt publication.

To the author in search of a publisher

Unpublished authors, especially, will find this booklet valuable and informative. **For your free copy, write to:**
**VANTAGE PRESS, Inc. Dept. F-53**
**516 W. 34 St., New York, N.Y. 10001**

WE STILL ARE
The first step in step soldering.

You take a giant step forward when you select solders from Indium Corporation for your step soldering operations. We offer solders with melting temperatures from 100° C to 363° C so you can solder at progressively lower temperatures without disturbing previously soldered electronic components and joints. One of the most popular solder systems for this purpose is our indium-lead system.

Please call for assistance in applications development or to order indium solders in preformed shapes, wire, ribbon, powder, or in one of our advanced paste formulations.

INDIUM CORPORATION OF AMERICA
Corporate Headquarters: USA • 800-4 INDIUM or 315-768-6400.
Indium Corporation of Europe: UK • (0234) 840255

can sense will tell you just where you are. You do need to keep strict books, nowadays digitally, and carry a good clock. Some input knowledge is indispensable: initial position and velocity as well as an adequate worldwide model of the pervasive pull of gravity. (George Gamow, genuinely an eminent physicist, ridiculed the task in Air Force circles around 1950 as one forbidden by general relativity. No one could accuse Gamow of stodgy reliability; his quicksilver mind had somehow overlooked the special simplicity of a gravitational field that is well known in advance.)

Navigating boxes guided most of the 1,300 V-2s fired in 1944 and 1945 against "the sprawling conurbation" of London, most big jets on long flights, and all hundred modified Scuds lofted out of Iraq in 1991. This volume is a study in detail of how it was all realized. Indeed, its highest realization remains in some doubt, for of course ICBM testing is partial and inferential. Only one U.S. missile warhead has ever gone all the way to detonation after reentry, a 1,200-mile Polaris test in 1962. The *Hubble Space Telescope* is a reminder of the precariousness of partial testing. There is thus room for doubt about missile accuracy, as about target hardness, but that doubt fades before other formidable hazards of nuclear counterforce.

On one page here are plotted the clustered impact points of 17 MX test missiles (no nuclear explosions, of course) sent 6,000 miles from California to splash into a Kwajalein lagoon in the mid-1980s. Analysis allows that cluster size to be inferred from the sanitized public testimony of an Air Force expert witness before Congress. The points do suggest some bias in the aim, but a bull's-eye 300 yards across still easily catches half of the impacts. For comparison, the commercial black boxes of the 747 would drift off course up to 10 miles on such a flight; the V-2 and the Scud, two ballistic missiles of comparable accuracy, might miss the atoll entirely by 50 miles and more, as extrapolated—for of course they have no such long range. Recall that the damage at impact point has risen from the V-2, able to smash houses over one city block, to the yawning crater a city block *deep* foreseen at each of the 10 impact sites of hydra-headed MX, to say nothing of the soft target alternative, 40 square miles of "smoking, radioactive" city per warhead.

Wherefore accuracy? The short answer is that it opens a new strategic stance in the time of nuclear plenty. Missile accuracy allows chesslike games of preemptive counterforce against small, deeply buried military targets. (Navy nuclear submariners, maybe predisposed by Alfred T. Mahan's old doctrine of blockade, eventually caught their bad case of counterforce fever from the Strategic Air Command.) The long answer is this compelling book, based on a first-rate nonmathematical exposition of the technology, a wide look at its history, and 140 interviews with listed political, military, engineering and intelligence experts, all of them participants. They represent those who formed and promoted the ideas, built the prototypes, organized the big instrument shops, found military contracts or sought the low-end civil markets, and watched over their Soviet counterparts at telemetry distance.

It should come as no surprise to any who have seen the expressive physicist Edward Teller on television how grandly a charismatic engineer leader—in this case, the late Charles Stark Draper of M.I.T.—could dominate the field of guidance. Draper too was an authentic technical pioneer, proponent of the floated single-degree-of-freedom integrating gyroscope, from his 1940s antiaircraft gunsight to the "beryllium baby" now whirring in all the MXs underground. (The Charles Stark Draper Laboratory is a not-for-profit R&D firm, spun off from M.I.T.; it produces no operational systems, taking contracts for prototypes only.) Three generations of Draper Lab guidance systems have improved in intrinsic, perhaps even operational, accuracy by five orders of magnitude. Time and money have redeemed the advocate's faith imbued by the engaging and confident Doc Draper to many an influential friend. A simple handshake with NASA chief Jim Webb "in a bar...in Germany" got Draper his first funds for the third generation of gyros, Doc used to say.

Long ago an international "gyro culture" had nucleated around a few physicists like Léon Foucault in the 1850s. By the 1930s two specialized firms shared most of the world military market: Sperry Gyro in the U.S. and Kreiselgeräte in the Third Reich. That quasi-governmental German firm provided guidance for most of the V-2s (some were radio-steered). The German experts brought to Redstone Arsenal after World War II continued with Kreiselgeräte developments for NASA and the U.S. Army. (Webb had in war-time been the treasurer of Sperry Gyro.) Gyros have long been widely used in aviation instruments; the best commercial gyrocompass at present is a Hungarian one, favored by the Soviets.

Guidance depends on more than money and sophisticated design: high craftsmanship. A 1977 Draper Lab report put it visibly: "No computer ever made a piece of working hardware. No software ever made a measurement, ground a fitting, sealed a vacuum, cast a bearing, wound a core...."

The most accurate Soviet guidance systems, "in the same league as the best Western ones," have shown the same incremental approach as have those of the U.S. The Soviet designs move in the path of Kreiselgeräte, using externally pressurized gas-bearing gyros, dry as a bone, with none of the fluorocarbon fluids that now buoy up each "beryllium baby." (Ball bearings are pretty passé.) Stark Draper fought gas bearings "like mad," but the other main actor in U.S. military inertial guidance, the Autonetics division of North American Aviation (now Rockwell International), has often included gas bearings in its more eclectic designs. At its zenith, theirs was no mean industry; during the years that Autonetics was building the guidance for 1,000 Minutemen, the firm employed 36,000 workers. Yet guidance is only a tithe on overall ICBM costs.

The larger claim of this fine book is hard to fault. No technology is external to its society, just as no society is external to its technology. Coupled to the world, technology cannot follow a literally inevitable trajectory. Yet convergence suggests that, given such sufficient conditions as a "state with an eye to possible war and as rich as the United States," it becomes likely that one or another of a family of relevant technologies will in fact be developed. Certainly any single path must first be made credible.

Establishing that credibility was "the key task" of the savvy political and technical leaders of inertial navigation. Perhaps stealthy jet bombers or cruise missiles would make plausible strategic weapons, as surely radio beams, star- and satellite-sighting and homing reentry are plausible routes to accuracy, like laser gyros. Would it much matter to the real world which of those had found its Stark Draper, its dazzling learning curve and its subsequent large-scale deployment?

To this reader, the direction of U.S.A.F. seems to have been broadly set, gyros or no, at least since Billy Mitchell in World War I argued for his airborne strategy. Dr. MacKenzie, a brilliant member of the Franco-Scots "sociology of science" mafia, may here be proving rather too much. Readers will concur in his concluding lines: "The time has now surely come to think the other unthinkable," and render us permanently safe from nuclear war.

# To preserve your copies of SCIENTIFIC AMERICAN

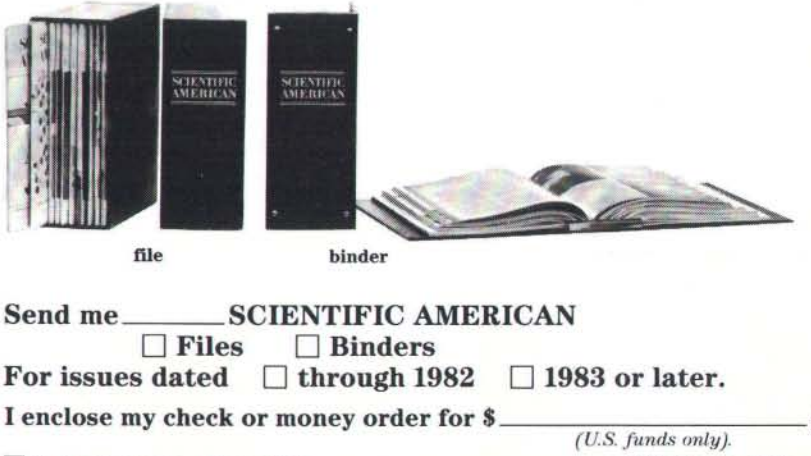
**A choice of handsome and durable library files or binders. Both styles bound in dark green library fabric stamped in gold leaf.**

**Files** **Each holds 12 issues. Price per file $7.95 • three for $21.95 • six for $39.95**

**Binders** **Each holds 12 issues. Issues open flat. Price per binder $9.95 • three for $27.95 • six for $52.95**

*(Add $1.00 per unit for postage and handling in the U.S. Add $2.50 per unit outside the U.S.)*

**To:** ***Jesse Jones Industries, Dept. SA, 499 East Erie Ave.,***
***Philadelphia, PA 19134***



**file** **binder**

**Send me______ SCIENTIFIC AMERICAN**
**☐ Files ☐ Binders**
**For issues dated ☐ through 1982 ☐ 1983 or later.**

**I enclose my check or money order for $____________________;**
*(U.S. funds only).*

☐ Charge my credit card $________________ (Minimum $15)
☐ Amex ☐ VISA ☐ MasterCard ☐ Diners

Card #____________________ Exp. date__________

Signature______________________________

**or call toll-free 7 days, 24 hours, 1-800-825-6690**

Name______________________________
*(please print)*

Address______________________________
*(No P.O. Box please)*

City______________________________

State____________________ Zip__________

*SATISFACTION GUARANTEED. Pennsylvania residents add 6% sales tax.*

# ESSAY: THE POETRY OF SCIENCE

by John Timpane

Early in this century, an old and great Dilemma resurfaced—Poetry or Science? It is truly old: Socrates was kicking poets out of his republic 2,400 years ago. It is truly great, for it is a fundamental problem in our culture, and it will not go away. Philosophers—John Dewey was one—felt that reconciling the scientific and the aesthetic was the most "significant question before the world." But could the two be reconciled? Which would best capture the essence of the age? Did the rise of one mean the decline of the other?

It looked bad for poetry. It was pre-Depression, pre-Hiroshima. Technology was the great success story; most Americans, poets included, tended to confuse it with "science." (A generation earlier, in the smokestacks, girders, trusses and machines of the Industrial Revolution, Walt Whitman had read the American soul: "Hurrah for positive science! long live exact demonstration!")

In truth, 20th-century poets owed much to science. In giving poets more things to think about, more viewpoints to take, science enlarged the ambit of the imagination and bade the poet take over from there. Planck's discovery of the quantum in 1900, followed by Einstein's theory of relativity in 1905, revealed a universe contrary to old assumptions and full of new points of view. Freud's work promised a new science of the mind. Anthropology revised history and Western humanity's worldview. Without these discoveries, without the new, relative universe they seemed to portend, no poet could have written, "Twenty men crossing a bridge, / Into a village, / Are twenty men crossing twenty bridges, / Into twenty villages," as Wallace Stevens did in 1918. And two acts of the imagination characterize the 1920s: T. S. Eliot's *The Waste Land* (1922) and Heisenberg's uncertainty principle (1927).

Science had something poets craved: the Scientific Gaze, the open-eyed, open-minded, skeptical point of view, dilated for the light of truth. William Carlos Williams aspired to such a gaze in his poem "Young Sycamore," which presumes to be nothing but sycamore: "dividing and waning / sending out / young branches on / all sides— / hung with cocoons— / it thins / till nothing is left of it / but two / eccentric knotted / twigs / bending forward / hornlike at the top." Like many other poets, Williams—whose studies led him to become a doctor—wanted poetry stripped of the "poetic" in favor of a lucid, unmediated truthfulness.

But Williams was a worrier; as much as he dwelled with science, he never came to terms. The mid-century poet who understood science best was Marianne Moore, a bohemian eccentric with a B.A. in biology from Bryn Mawr. Her scientific gaze was drawn to the quirky side of nature; in her surprising, angular poems about pangolins, baseball, ostriches and monkeys, a belief emerges that science and poetry are part of the same quest for illumination. What poetry gives us, Moore writes, is "imaginary gardens with real toads in them." That stunning, funny phrase means, among many other things, acts of imagination grounded in reality. Poetry, meet science.

Three decades later, post-Hiroshima, mid-cold war, the great Dilemma arose anew in the passionate "two cultures" debate. It took a scientist-poet who had seen Nagasaki—Jacob Bronowski—to apply the quietus, which he did in his books *Science and Human Values* and *The Visionary Eye*. Bronowski depicted culture as a unity in variety: "There is a likeness between the creative acts of the mind in art and in science." These two acts, different but related, can yield between them a fuller view of the cosmos. One of his best poems suggests that the principle of reality is "both abacus and rose combined."

Now, three decades later still, we can declare the Dilemma dead. There is no Dilemma and never has been. Socrates was wrong. These two great ways of seeing lie on the same imaginative continuum. They do not compete; they connect. If science explicates the surprising, complex, undreamed of truth, poetry enacts the full impact of that truth on the human consciousness. A good poet can take an insight revealed by science and suggest the full range of its human importance.

Poets writing at this very moment—some good ones are Gary Snyder, Albert Goldbarth, Elizabeth Socolow, Ed Dorn, A. R. Ammons and Pattiann Rogers—are doing just that. In "Corsons Inlet," by Ammons (who got his B.S. at Wake Forest College), the speaker observes the natural world and lets it teach him something: "in nature there are few sharp lines: there are areas of / primrose / more or less dispersed; / disorderly orders of bayberry; between the rows / of dunes, / irregular swamps of reeds... / I have reached no conclusions, have erected no boundaries, / shutting out and shutting in, separating inside / from outside: I have / drawn no lines." Ammons's poem is an exhilirating, harsh, open-ended world, full of beauty, pain and chaos—that which follows rules but cannot be predicted. Written in 1965 but full of chaos theory, "Corsons Inlet" is not the less "poetic," is indeed more beautiful, for invoking science.

The same is true for "Seduced by Ear Alone," by Pattiann Rogers. Where to start a love poem? With the physics and physiology of hearing: "Someone should explain how it happens, starting / With the dull stimulation of anvil and stirrups, / The established frequency of shifting air molecules / Initiated by your voice, entering my ear... / if the wind in the trees / Should sound by accident like the timbre of your voice, / I can be fooled for an instant, feeling suddenly in the dark / Estimable and saved." Here, science helps give a poem its humor and its humanity.

Near century's end, more poets know more science better than any poetic generation in history. The result is a poetry that seeks to situate the reader in the world in all that world's complexity. One simply cannot do that without knowing about quarks, Julia sets, angular unconformities, the equilibrium hypothesis of island biogeography, microeconomics, superconductivity, Saussurian and Chomskyan linguistics, the rings of Uranus—and every poem since *Gilgamesh*. Poets have their work cut out for them, and much of that work is science.

So? So read. Scientists, read poetry. Poets, read science. All will find themselves faced with a lifelong education. Do not expect easy access or overnight conversion. Realize that all stand to gain. In science, poets will gain the world and keep their souls; in poetry, scientists can gain entry to imaginary gardens for their real toads.

---

*JOHN TIMPANE is a writer who teaches English at Lafayette College.*

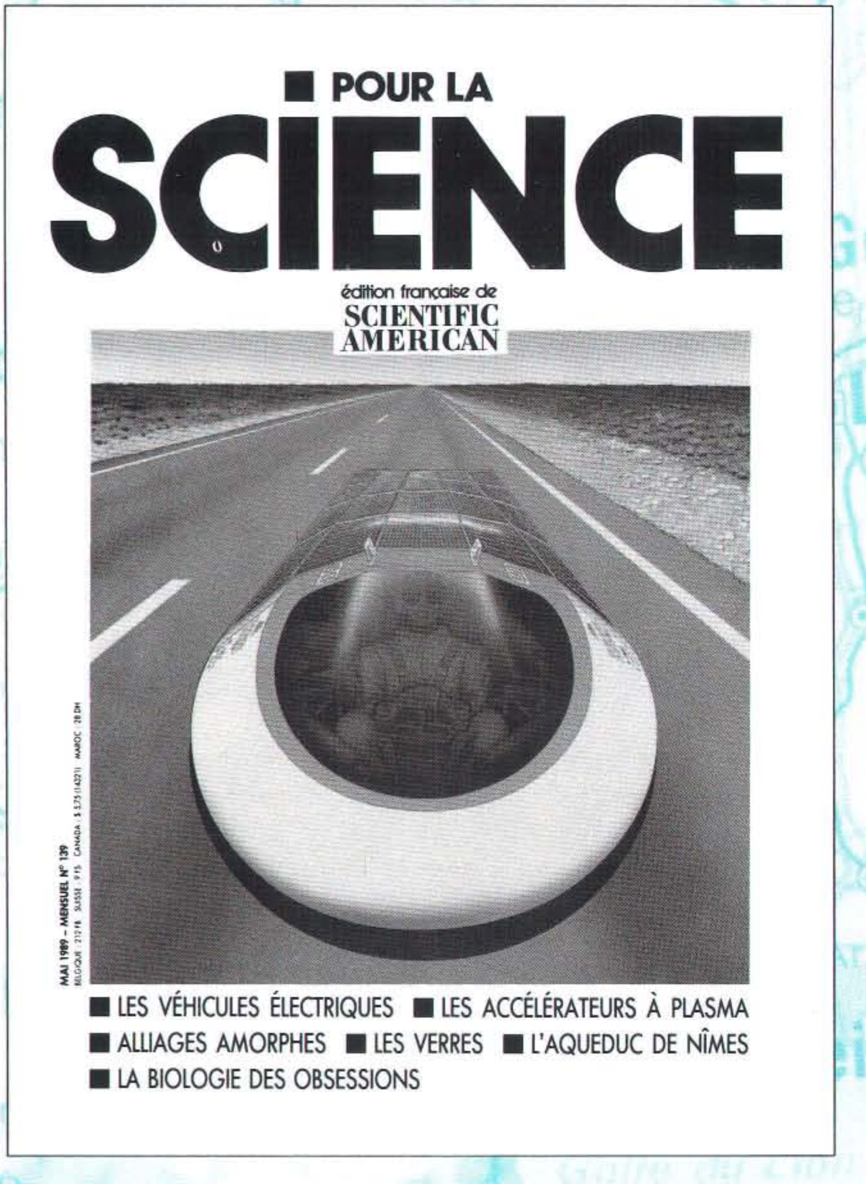
# Your business can reach the people who make the future happen in France.

**POUR LA SCIENCE** is the French-language edition of SCIENTIFIC AMERICAN. More than 60,000 people buy this prestigious science and technology magazine each month, people who make decisions for France's industry, government and business sectors.

Contact
Susan Mackie at:

**POUR LA SCIENCE**
8, rue Férou
75006 Paris, France

**Telephone**
(33) (1) 46-34-21-42
**Fax**
(33) (1) 43-25-18-29
**Telex**
842202978



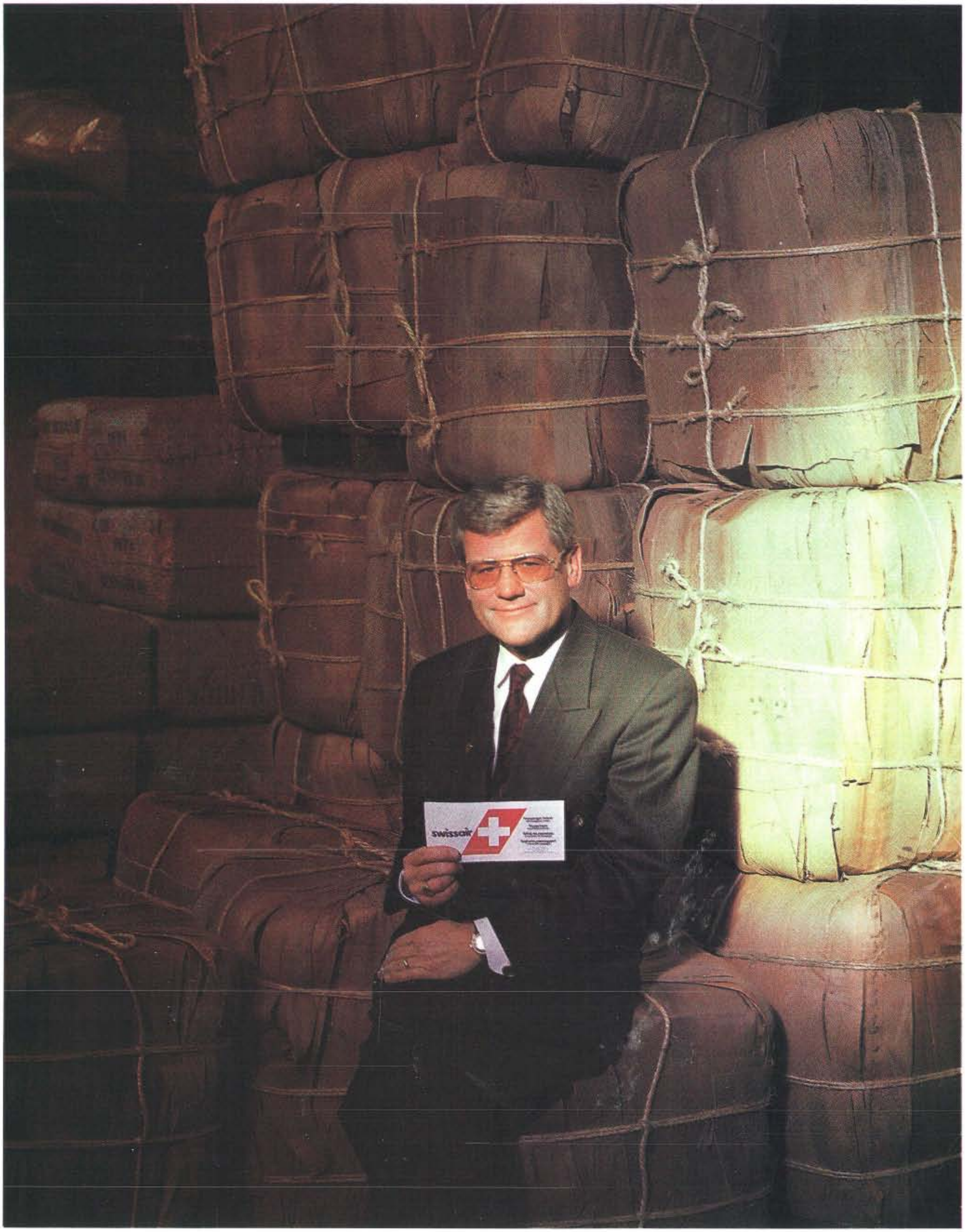
POUR LA SCIENCE

édition française de SCIENTIFIC AMERICAN

MAI 1989 – MENSUEL N° 139

- LES VÉHICULES ÉLECTRIQUES
- LES ACCÉLÉRATEURS À PLASMA
- ALLIAGES AMORPHES
- LES VERRES
- L'AQUEDUC DE NÎMES
- LA BIOLOGIE DES OBSESSIONS

Swissair Customer Portrait 130: Karel Van Susante, entrepreneur in the tobacco industry, Kampen, Netherlands.



Only the very best is good enough for him. From selecting his tobacco blends to choosing an airline. So when he flies off for a personal assessment of the quality of the plantations, the outstanding comfort of the all-new MD-11 jet is a further reason for his choosing Swissair. The natural choice of someone who makes his selections with the greatest care.